Lecture Notes in Medical Informatics

Edited by D. A. B. Lindberg and P. L. Reichertz

4

Clinical Trials in 'Early' Breast Cancer

Methodological and Clinical Aspects
of Treatment Comparisons
Proceedings of a Symposium, Heidelberg, Germany,
4th to 8th December, 1978

Edited by
H. R. Scheurlen, G. Weckesser and I. Armbruster

Springer-Verlag
Berlin Heidelberg New York 1979

Editors of this Volume

H. R. Scheurlen
G. Weckesser
I. Armbruster

Institut für Medizinische
Dokumentation, Statistik
und Datenverarbeitung
der Universität
Im Neuenheimer Feld 325
D-6900 Heidelberg 1

ISBN-13:978-3-540-09530-9 e-ISBN-13:978-3-642-93118-5
DOI: 10.1007/978-3-642-93118-5

Library of Congress Cataloging in Publication Data
Main entry under title:
Clinical trials in 'early' breast cancer.
(Lecture notes in medical informatics; v. 4)
Bibliography: p.
Includes index.
1. Breast--Cancer--Prognosis--Statistical methods--Congresses.
I. Scheurlen, Hans. II. Weckesser, G., 1948- III. Armbruster, Inge, 1953- IV. Series.
RC280.B8C59 616.9'94'49 79-19183
ISBN-13:978-3-540-09530-9

2141 /3140-543210

P r e f a c e

In the present situation of clinical trials it seems a worthwhile task to bring clinicians and statisticians together to talk about common problems. When, in summer 1978, the Deutsche Forschungsgemeinschaft made available the resources for a scientific meeting we did not hesitate to submit such a cooperative project.

We are grateful to the members of the Sonderforschungsbereich 123 'Stochastische Mathematische Modelle' for giving precedence to that project.

Above all we thank Prof. Dr. H. Immich who gave us support with word and deed from the very beginning.

Our thanks are also due to Sarah Nelson who helped us by looking through the comments as well as preparing the papers for publication.

Finally we would like to thank Mrs. Heidrun Wunsch for her expert re-typing most of the papers.

Following the idea of our symposium much attention is given in this book to discussions. Thus the reader may form a picture of whether such a meeting deserves to be repeated.

H.R. Scheurlen
G. Weckesser
I. Armbruster

CONTENTS

I n t r o d u c t i o n

Hans Scheurlen
Institut für Medizinische Dokumentation,
Statistik und Datenverarbeitung
Universität Heidelberg

Thirty years ago Paterson and Russell initiated the first controlled clinical trial in the field of treatment for breast cancer. At almost the same time Boag published his paper on "maximum likelihood estimates of the portion of patients cured by cancer therapy". Since then, both clinicians and statisticians have been increasingly devoting themselves to the problem of how to express the efficiency of treatments in terms of the patients' chance of survival. The incorporation of regression-like arguments into life table analysis, as proposed by D.R. Cox in 1972, seems to apply to a vast variety of real situations and has led to a remarkable number of papers from the statisticians.

However, we also have reason to view with concern what is going on. Do the new concepts inspire the clinical experimenters to put their problem more precisely? Do controlled trials have any influence on the medical practice of treating patients at all?

We have addressed these questions in a more detailed form to expert clinicians and statisticians (see the appendix of the Discussion), inviting them to discuss these questions on the occasion of our symposium. The limitation of the general theme to a particular stage of a particular disease served two purposes: the selection of the clinicians and the stimulation of the participants to reach some feasible results.

To give a provisional answer I feel that the clinicians and the statisticians have become progressively estranged, this being the other side of the recent pleasing productivity of the statisticians. If that is the case I think it is unpleasant. It may be useful in this context to consider the medical history (I) and the present state (II) of our problem.

I

In the nineteenth century, during the period of antisepsis, French and German surgeons had a violent controversy about operations in the case of breast cancer. Verneuille, a spokesman of the French, objected to

operating on patients unless the disease seemed to be in a very early stage. In the vast majority of cases, however, he considered breast cancer to be far advanced and thus any operation to be useless or even harmful, since it frequently was seen to be followed by a dramatic turn to the worse. Billroth from Vienna, the leader of the German surgeons, wanted to give operations in all cases. He maintained that a considerable number of patients was spared in this way and that some patients even lived for many years without recurrence. His failures are nevertheless indisputable, as one can see from the case reports, published in 1878 by A. von Winiwarter.

William Stewart Halsted had the opportunity to watch Billroth working when he visited Vienna in 1879. From that he concluded that the technique used in Vienna was without any sound foundation, leaving parts of the tumour mass unremoved in most cases. As Haagensen pointed out, Billroth in fact was "treating breast cancer with what we would today describe a simple mastectomy and in some patients also with limited axillary dissection". Stressing the locally advanced but non-systemic nature of the disease, Halsted consequently developed his radical technique of mastectomy when he returned to New York.

The short-term results of the radical mastectomy looked quite promissing. Some ten years later, however, surgeons had to confess that despite a marked decrease in local recurrence rates the majority of patients died with metastatic disease anyway. In 1902 Pusey, Beck and Turner reported on the radiosensitivity of inoperable breast tumours and described some successful cures using radiotherapy. Some doctors therefore hoped to improve their results by giving radiotherapy in addition to radical mastectomy. In Germany some hospitals began to routinely use postoperative radiotherapy in 1912. As a result a further decrease of local recurrences was found by those using relatively high radiation doses. The question whether adjuvant radiotherapy has an influence on the survival time as well caused a dispute between surgeons and radiologists. Papers concerning this controversy caused the literature to swell like an avalanche.

During the thirties, however, critical contemporaries already had their doubts about the validity of the "Halsted doctrine". If the prevention of local recurrences is not related to some substantial improvements of the patients' chance of survival then the disease cannot be limited at the time of operation so frequently, as was assumed previously. McWhirter considered the surgical manipulation itself a possible cause of dissemination. He proposed the hypothesis, that the frequency of

such an event is related to the extension of a manipulation anyway. In 1941 more than forty surgeons from Southeast Scotland decided to replace the classic Halsted procedure by a combination of simple mastectomy and postoperative irradiation. As far as possible all women with breast cancer from that region were included. By a historical comparison McWhirter, in 1949, drew the conclusion that the so-called McWhirter-method is superior to radical mastectomy (with or without adjuvant radiotherapy).

Using historical comparisons is a dangerous procedure because it is impossible to distinguish between a real difference in treatments and (what Berkson called) a general "time trend" towards a more favourable survivorship of primary treated patients. As Berkson has pointed out this "time trend" does not vanish even if the samples are stratified for comparison by aid of relevant prognostic factors. Moreover, the question whether unrandomized series are comparable, whatever their nature, remains "the devil's own question" being the root of all confusion within the medical literature up to now.

Meanwhile the physicians have perceived that randomization of patients is a necessary prerequisite for treatment comparisons. Since the onset of the Manchester trial thirty years ago a lot of clinical trials have been carried out and analysed (whether intermediately or finally). Physicians are now in a position to judge the situation more cautiously. Unfortunately they do not yet agree, even on those points which have been re-examined repeatedly in the past. Recently clinical trials have been aparently focussed mainly on two problems:

(i) Which is most effective, a conservative or radical operation?
(ii) Is there any improvement of the results of local therapy when adjuvant systemic treatment was added?

II

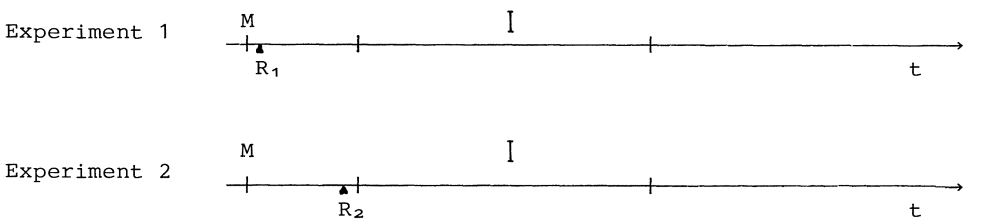
The motives for planning an experiment are greatly different, comprising conflicts with traditional philosophies (such as the "Halsted doctrine") as well as simple chance discoveries. However, it seems impossible to find any experiment to be derived from others in a more systematic manner. Whenever the clinician is consulted by a patient he will be faced with a complex variety of treatment policies which cannot be kept on a simple conservative-radical scale. However, a comparison of more than two or three treatments at once is very rare amongst the reported trials.

Are there any common conclusions therefore to be drawn from the results of several experiments when the treatments of one experiment are, partly or completely, different from those of another one, or when neither the criteria for acceptability of patients nor the criteria for stratification are comparable? What conditions are necessary for experimental strategies to be comparable enough to make common conclusions?

Instead of giving an answer I would like to stress another point which is connected with the question of compatibility and with some special features of clinical experimentation as well. When devising and performing clinical trials we must be aware of the fact that patients on study are entitled not only to the best possible therapy but also to the freedom of choice whether to continue or to refuse the treatment assigned to them by random allocation. If this is true, protocol deviations will be inevitable. The subsample comprising those deviants is by no means randomly drawn from the population of the accepted patients. Its size and bias will have an influence on the result of the experiment. Size and bias themselves are related to some experimental conditions such as

(i) the patients' "informed consent",
(ii) the mental and somatic stress due to the treatments,
(iii) the time of randomization.

To make the third point clear let us consider two experiments with postoperative irradiation being one of the treatments and let the time of randomization be the only difference between the two experiments, this being soon after the mastectomy (at R_1) and right before the beginning of irradiation (at R_2) respectively (see the following figure with M, R_1, R_2 the time of mastectomy and randomization respectively and I the period of irradiation).



Routinely used irradiation is impossible when patients have a relapse between R_1 and R_2. Patients possibly refuse irradiation when they are prior to R_2 either in an excellent condition or a very poor condition from causes other than cancer. In experiment 1 all these patients are

protocol deviants causing bias when excluded from evaluation and leading to ill-defined treatments when put on test. In experiment 2 handling this sort of patients is no problem at all. However, in experiment 2 the target population is not quite the same as in experiment 1.

The clinician is finally interested in the "significance" of the result of his experiment. He is now faced with questions concerning the choice of time for testing and the choice of test procedure to be applied, giving rise to some arbitrariness. For example let us assume that a decision is to be taken between a conservative policy and a radical one. Let us further assume that the portion of patients cured as well as the portion injured is raised by the radical treatment. As a consequence some patients will die either very early or very late after the radical treatment as compared with the conservative policy. Possibly the result of a trial might be "not significant" when using a plausible procedure such as the generalized Wilcoxon test but "significant" when using the logrank test which looks just as plausible. What is the meaning of "significant"?

Clearly the answer depends on what the clinician really wants to find out, i.e. on the statement of his problem and the formulation of his hypotheses. Some subjects of those statements are: the probability that a patient survives x years, the life expectancy, the median survival time, the portion of patients cured, some net or partial crude probabilities, some parameters of a particular distribution function, the proportionality of hazard rates, the interaction between some treatment components or between treatments and some concomitant variables, the doubling time, the risk of nodal involvement or the risk of remote metastases within some specified interval. To answer the question as to whether one treatment is preferable to another one it may happen that

(i) different answers are given according to how the question was specified (and what model was adopted);

(ii) a difference of treatment effects, even though substantial, is hard to interpret in terms of a "better than" relation;

(iii) a significant difference of treatment effects is undetectable because some internal observations, such as response, were not available.

I will not enter now into a debate of principle on comparative statistical inference. I would rather want you to bear in mind that we must not describe the result of a trial in terms of simple statements of significance levels or interval estimates. Treatment effects are not

only hard to describe but also likely to be rather small. As a consequence an increasing proportion of diseased women will have to be treated under the conditions of an experiment. We must therefore maximize the amount of information which can be obtained from a clinical trial. This can only be done by a clear formulation of the hypotheses on which the trial is based, right from its conception. Clinicians and statisticians must be brought together in planning the type and number of measurements needed to test these hypotheses. Precise definitions of the criteria of assessment are needed, together with a combined understanding of the practical value of statements about the "significance" of a treatment difference as measured by a particular parameter.

I feel that in the future we will have to concentrate our attention on both the organizational and methodological problems of the coordination of outstanding trials.

Prognostic Factors and Nosological Criteria of Breast Cancer from the Pathologist's Point of View

R. Baessler

Pathologisches Institut der Städtischen Klinik
Fulda, Federal Republic of Germany

Patho-morphological examination of the surgical specimen yields some reliable data which are of a considerable prognostic relevance. They are essential to the clinician, planning therapy, and to the statistician as well when he is occupied with the analysis of treatment effects using stratification. It should therefore be useful to consider how the pathologist proceeds in producing those data.

(1) Size and Form of the Primary Tumour

We start the examination of a specimen with the inspection and palpation of the cut surfaces using the eyes or sometimes radiographs (with appropriate labelling) and the fingers. Microcalcifications if present will be excised carefully by aid of simple radiographs.

The mastectomy specimens are cut into parallel sections and the cut surfaces evaluated macroscopically. In well delimited tumour the size can be readily determined. In diffusely growing carcinomas this is, however, difficult or even impossible. On the other hand, a hematoma cavity may simulate a much larger carcinoma than is really present.

Determination of tumour size is thus not always possible and sometimes uncertain. We found during our own investigation on about 3.500 specimens in the period 1961 to 1971 that before therapy 40% of the carcinomas had a diameter up to 3 cm and only 20% with a diameter smaller than 2 cm (Fig. 1). As compared with our last figures there is an increase of small carcinomas (20 to 28%) and a decrease of large tumours i.e. more than 4 cm in diameter, (33 to 23%). On an average the diameters decreased from 3.7 to 2.5 cm. The long-term prognosis in small tumours is accordingly improving over time.

Determination of the size of the primary tumour is highly significant since it is closely related to the frequency of lymph node metastases. We found about 20% axillary metastases when the diameter of the primary was 1 cm. With increasing tumour size, the proportion of diseased lymph nodes increases to 30% at 3 cm and to 50% at 5 cm. Following HAAGENSEN (1971) the overall frequency of axillary metastases in some

carefully documented modern series of radical mastectomy amounted to 50%.

The form fo the carcinomas is of some prognostic relevance too. According to LANE et al. (1961) and to GALLAGHER and MARTIN (1969, 1971), one can distinguish macroscopically between stellate or polygonal and round or nodular carcinomas (Fig. 2). The stellate form is strongly related to invasive ductal carcinomas whereas the round form is related to differentiated or medullary types having a much more favourable prognosis.

(2) Patho-histological Classification

The histo-pathology of tumours serves for characterization, identification and histo-genetic classification. Its prognostic significance, although still in dispute, was stressed only recently. According to our own investigations, we classify the carcinomas into two groups similarly to the WHO. It is not possible here to discuss the individual groups systematically. I would, however, like to consider a few points.

Group 1: The invasive ductal or non-differentiated carcinomas (previously often referred to as carcinoma simplex, infiltrating carcinoma, carcinoma not otherwise specified) axhibit a pattern of various solid and adenomatous cells. All results of epidemiology, therapy and prognosis are measured against this type of carcinoma. The tumour cells are usually arranged in cords, nests and glandlike structures. The WHO subdivides this category into two groups on the basis of the amount of fibrosis within the tumour as "invasive ductal carcinoma with little or marked fibrosis".

Group 2: The intraductal carcinoma is characterized by a direction of growth along the major mammary ducts. Encroaching on the glandular lobules it possibly induces a secondary or transmitted lobular carcinoma (Fig. 3). It also may encroach on the nipple giving rise to the diagnosis of Paget's disease. Immunofluorescent microscopic studies have shown that the cells of Paget's disease are always of ductal origin.

These carcinomas give rise to difficulties in making diagnosis in two respects:

a) It is known that invasive ductal carcinomas often have an intraductal growing component, provided they are undifferentiated (Fig. 4). So the classification of the tumour may be ambiguous.

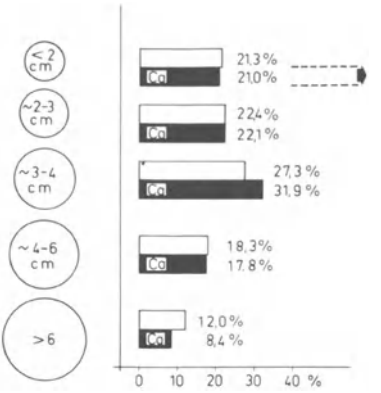


Figure 1 Average size of breast tumours of 3464 specimens: Black columns indicate the diameters of cancers, white columns benign tumours and dysplasias.

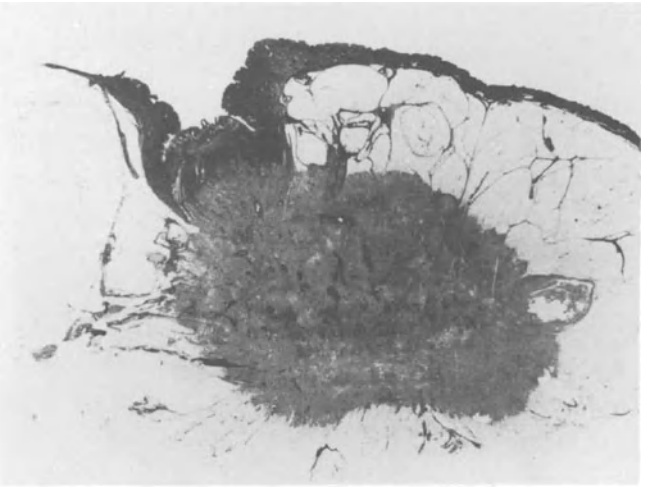


Figure 2 Gross-section of a well-delimited large breast cancer of 7 cm in diameter with retraction of the mamilla. Histologically an invasive ductal carcinoma.

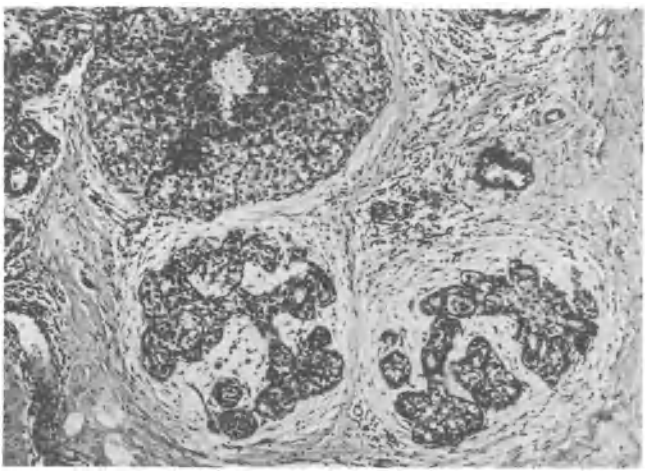


Figure 3 Intraductal carcinoma with encroaching the glandular lobules, i.e. a secondary lobular carcinoma. Magnif. 140 x

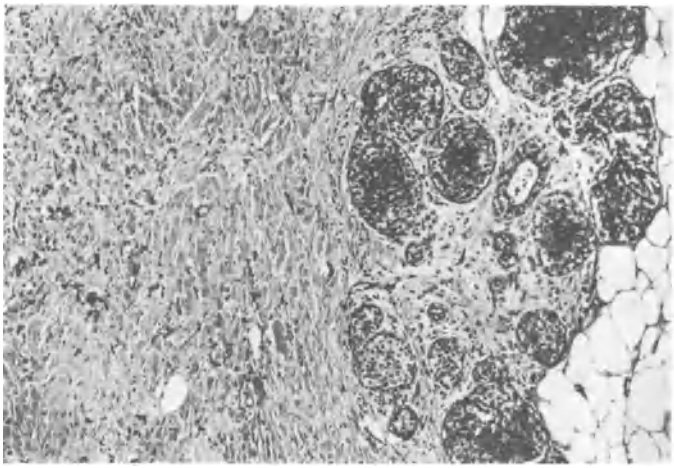


Figure 4 Invasive ductal carcinoma with development of an intraductal component of the cancer, infiltrating the fat tissue in the right part of the picture. Magnif. 140 x

b) Since the proliferation of the duct system might be kept within natural limits for a long time, pre-invasive phases with different grading are observed particularly in these carcinomas (Fig. 5). Between 75% and 90% of carcinomas which had their origin within the duct must be diagnosed as invasive tumours (Fig. 6). This means that the non-invasive phase is an exception and must be proved.

Group 3: This group comprises a mixture of carcinomas which are well-distinguishable as to the nature of their differentiation:

a) the mucinous carcinoma;
b) the medullary carcinoma with lymphoid stroma and connective tissue capsule, the so-called circumscribed carcinoma;
c) the adeno-cystic carcinoma;
d) the squamous carcinoma and some other rare forms.

Group 4: The lobular carcinoma appears to be a proliferation of the solid epithelium bounded by the natural limits of the lobule and the terminal ducts. The cells are largely isomorphic and the chromatin content is raised. According to our own studies, these tumours derive from the basal cells of the lobular epithelium which are capable of developing intracytoplasmatic fibrils.

Now I would like to say a few words on the carcinoma lobulare in situ and its prospective potency of changing into an invasive growth. The time elapsing from onset of a carcinoma in situ to the beginning of an invasive growth is clearly unobservable. We can only observe the time interval between the previous diagnosis of the carcinoma in situ and the subsequent diagnosis of an invasive small cell carcinoma. The observed intervals range from a few to twenty or thirty years. HUTTER and FOOTE (1969), HAAGENSEN et al. (1972) and ANDERSEN (1974) have reported on a total of 136 patients having had a biopsy out of the contralateral breast within 2 to 28 years subsequent to diagnosis of the carcinoma in situ. An invasive carcinoma was found in 13 cases (9.5%). One hundred patients also had a biopsy out of the ipsilateral breast. From these patients an invasive carcinoma was found in 23 cases (23%).

(3) Cellular Defense Reaction

We know that a particular proportion of the carcinomas does not display any cell reaction in the periphery i.e. the boundary zone of the stroma. Lympho-plasmacellular infiltrates in the periphery are on the other hand a typical finding in the case of medullary carcinoma(Fig.7).

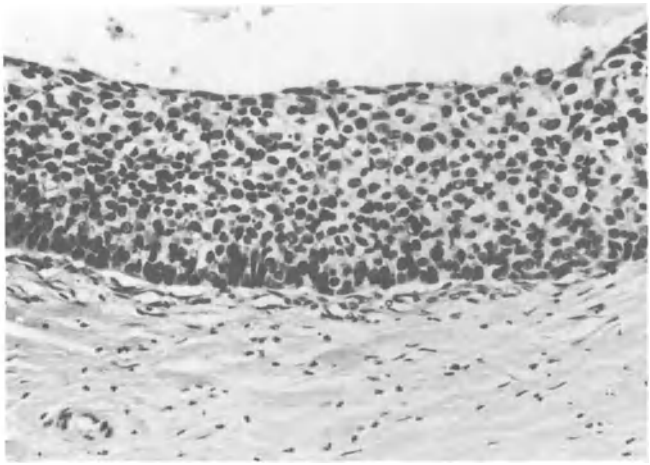


Figure 5 Prospective development of the lobular carcinoma in situ. Correlations of the frequency (%) and the time interval by several authors (By Baessler, 1978)

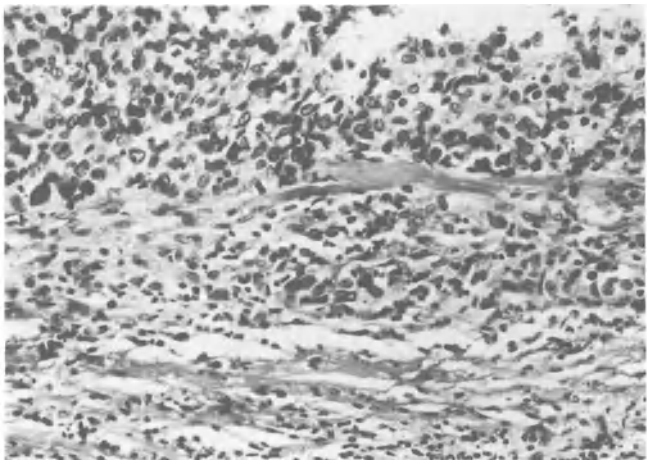


Figure 6 Non-invasive intraductal carcinoma with irregular and hyperchromativ nuclei (high grade type). Manif. 180 x

According to our own studies (and that of FISHER et al. (1975)) there are

no cellular infiltrates	in 19 (24) %
slight or moderate infiltrates	in 62 (58) %
strong cellular infiltrates	in 19 (17) %

(the figures of Fisher are in brackets).

Results up to now (BLACK et al. (1971), BERG (1971), HAMLIN (1968) and others) can be summarized as follows: a cellular reaction within and around the carcinoma indicates a defense reaction (Fig. 8). Its intensity as well as the composition of the infiltrates depend on the antigenicity of the tumour and the degree of malignancy.

In the cellular infiltrates the lymphocytes predominate with 46%. According to SCHOORL (1976) these are T-lymphocytes manifesting a cell-mediated immunity. B-lymphocytes predominated in intraductal carcinomas.

(4) Tumour Grading

Let us now turn to the methods of estimating the degree of malignancy. Any concept of a rank order of malignancy must take into account
(i) the resultant of the aggressive behaviour of the tumour,
(ii) the resistance mechanisms of the host, and
(iii) the differentiation and dedifferentiation of the tumour.

In evaluating breast cancer we, as most of the other pathologists today, refer to the system of BLOOM and RICHARDSON (1957). Other criteria are preferred by HARTVEIT (1971). The histological criteria for grading are:
(i) degree of differentiation as to the formation of tubuli and acini,
(ii) pleomorphism of cell nuclei,
(iii) hyperchromasia and mitoses.

Usually we subdivide the findings into three groups according to a scoring system:
Grade I corresponds to a low degree of malignancy.
Grade II corresponds to an intermediate degree of malignancy.
Grade III corresponds to a high degree of malignancy.

Frequently, however, the agreement among pathologists, when considering the same case, is poor, despite careful examination. Above all grade II shows large fluctuations and grade I differs between 8 and 26%, demonstrating the weakness of this method. Nevertheless, investigation series showed some correlation between tumour grading and the survival time. From about four thousand cases of different authors five-year survival rates were determined for grade I = 81%, grade II = 53%, grade III = 30%.

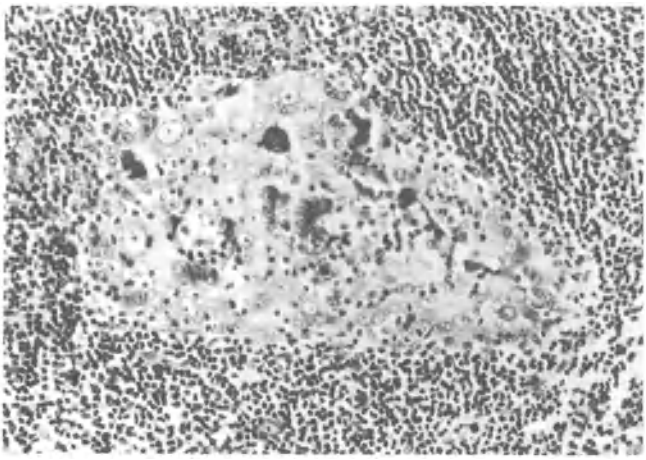


Figure 7 Medullary carcinoma with lymphoid stroma. Magnif. 230 x

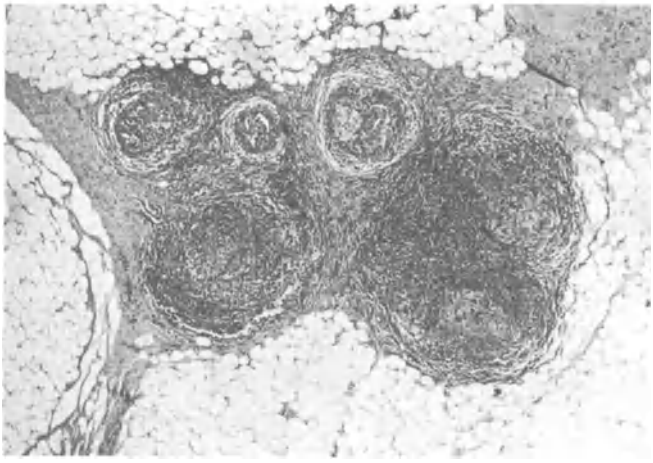


Figure 8 Strong defensw host reaction in the surrounding of a invasive ductal carcinoma with small tumour cell groups and predominate lymphocytic infiltrates. Magnif. 70 x

(5) Axillary Lymph Node Metastases

As a rule, the search for metastatic growth in axillary lymph nodes is based on the examination of seven to twelve nodes out of the content of the arm pit. By careful examination, 32 to 35 lymph nodes are detectable in this material. The mean diameter of uneffected lymph nodes is 6.5 mm whereas effected lymph nodes have diameters between 7.9 to 40 mm. Tumour nodes up to 2 mm are designated as micrometastases (macrometastases otherwise). Technically we proceed in such a way that the lymph nodes are extracted. If this was too difficult, then the lymph nodes could be shown up well by Bouin fixation.

As it was already stated above in the discussion of tumour size, there are close relations between the size of the primary and the frequency of axillary lymph node metastases. On the other hand there is no relationship between frequency of metastases and age.

The frequency distribution of the metastases according to the topological groups of regional lymph nodes are shown by the following figure (Fig. 9). We see that the central, interpectoral and the circumvenous lymph nodes are most frequently affected. Provided that only one metastasis is present, we usually find it in the central group.

When all regions are affected by the tumour, the 10-year survival is 30% (HAAGENSEN (1971)).

The source of axillary metastases may be situated anywhere within the breast whereas metastases of the sternal lymph nodes are more strongly related to cancers of the inner quadrants and the centre. However, 25% of sternal lymph node metastases have their origin also in the outer quadrants.

The Duration and Course of the Disease

Investigations by BLOOM (1965) on the natural history of untreated breast cancer, including 250 cases from the Middlesex Hospital out of total of about 1,000 compiled cases revealed the following: 95% of the women died of immediate consequences of the tumour condition, especially of metastases. 5% died of intercurrent diseases. The mean duration of symptoms until hospital admission was 3.25 years, the mean survival time was 2.7 years. 18% lived for five years and 4% for ten years. Spontaneous cures were not observed. Investigations on treated metastatic breast cancer (150 autopsied cases) reveals acute complications as the cause of death in the first two years after breast amputation, and sequelae of the progressive metastases as a rule in the third up to

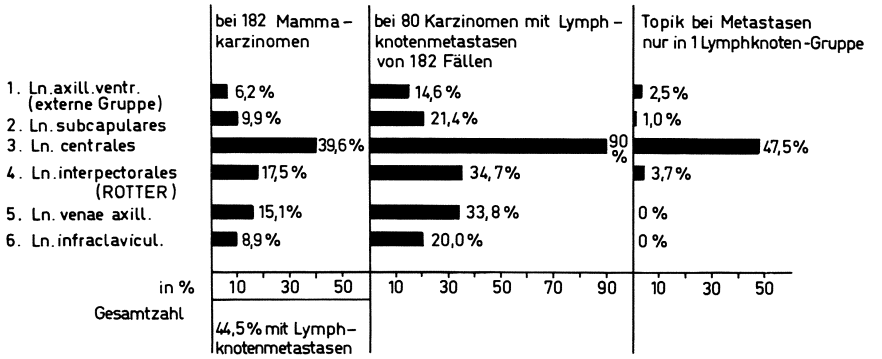


Figure 9 Frequency and the topical pattern of the metastases in the axillary nodes of breast cancer, with statistical dates by Haagensen (Baessler, 1978)

the fifth year.

The most frequent causes of death are: pulmonary embolism, anesthetic incidents, pulmonary and cardiac insufficiency in pulmonary metastases, mass hemorrhages, pericardial tamponade, electrolyte disorders and hepatic coma.

We were further interested in the question whether the pattern of metastases depends on the duration of the disease. We therefore divided our cases into two groups as to whether the survival time was equal to or less and more than 3.5 years respectively. We found no differences at all, despite the fact that some more recurrences of the scar and the skin occurred among the longer living patients.

References

Andersen, J.A.: Lobular carcinoma in situ: a histological study of 52 cases.
Acta path. microbiol.scand. A 82 735-741 (1974)

Baessler, R.: Pathologie der Brustdrüse, In:Doerr, W., Seifert, G., Uehlinger, E.: Spez. Pathologische Anatomie, Bd. 11. Berlin, Heidelberg, New York Springer 1978

Berg, J.W.: Morphological evidence for immune response to cancer.
Cancer 28, 1453-1456 (1971)

Black, M.M., Asire, A.J., Leis, H.P.: Cellular responses to autologous breast cancer tissue. Correlation with stage and lymphoreticuloendothelial reactive.
Cancer 28, 263-273 (1971)

Bloom, H.J.G.: The influence of delay on the natural history and prognosis of breast cancer.
Brit.J.Cancer 19, 228-262 (1965)

Bloom, H.J.G., Richardson, W.W.: Histology grading and prognosis in breast cancer. A study of 1409 cases of which 359 have been followed for 15 years.
Brit. J. Cancer 11, 359-377 (1957)

Fisher, E.R., Gregorio, R.M., Fisher, B.: The pathology of invasive breast cancer.
Cancer 36, 1-85 (1975)

Gallagher, H.S., Martin, J.E.: Early phases in the development of breast cancer.
Cancer 24, 1170-1178 (1969)

Gallagher, H.S.: An orientation to the concept of minimal breast cancer.
Cancer 28, 1505-1507 (1971)

Haagensen, C.D.: Diseases of the Breast. Philadelphia, London, Toronto. W.B. Saunders Co. 2nd ed. 1971

Haagensen, C.D., Lane, N., Lattes, R.: Neoplastic proliferation of the epithelium of the mammary lobules. Adenosis, lobular neoplasia, and small cell carcinoma.
Surg. Clin. N. Amer. 52, 497-524 (1972)

Hamlin, J.M.: Possible host resistance in carcinoma of the breast. A histological study.
Brit.J. Cancer 22, 383-401 (1968)

Hartveit, F.: Prognostic typing in breast cancer.
Brit. med.J. 4, 253-257 (1971)

Hutter, R.V.P., Foote, F.W.R.: Lobular carcinoma in situ. Long term follow up.
Cancer 24, 1081-1085 (1969)

Lane, N., Goksel, H., Salerno, R.A., Haagensen, C.D.: Clinicopathologic analysis of the surgical curability of breast cancers: A minimum ten-year study of a personal series.
Ann.Surg. 153, 483-498 (1961)

Schoorl, R., de la Riviere, A.B., von dem Borne, A.E.G., Feltkamp-Vroom T.M.: Identification of T and B lymphocytes in human breast cancer with immunhistochemical techniques.
Amer.J. Path. 84, 529-544 (1976)

Comment to Dr. Baessler's paper

Dr. Baum

There is little known about the natural history of untreated breast cancer. I have reviewed the evidence (Baum 1977). In summary, what little evidence there is, suggests that untreated early carcinoma of the breast could have the same prognosis as that of treated early carcinoma of the breast. There has been an error in the past to consider the studies of the natural history of advanced carcinoma of the breast as giving us information about the untreated disease as a whole. I think that is a false assumption.
With regard to extent of node sampling, Fisher's studies have demonstrated that it makes no difference how many nodes from the axilla are sampled if four are found to be involved; whether it is four out of

four or four out of thirtyfour nodes examined, the outcome is equally bad.

Ref. Baum 1977. The Curability of Breast Cancer.

Growth Rate of Tumours and Natural Life Expectancy

D.v.Fournier

Universitäts-Frauenklinik Heidelberg

Patients and Methods

With known volume doubling times growth rates and life spans of tumours are projectable. For the detection of 'early' cancer it is necessary to know: The growth time from inception of the cancer until it reaches a detectable size and the length of time between this threshold size and the time the tumour is no longer curable.

In 147 mammary carcinomas from various therapeutic centres all over the country 388 serial mammographies were performed before final treatment. Serial mammographies were done due to delay of the final diagnosis, the refusal of treatment and other reasons. The average observation time was 2,25 years with a range of 0,25 years to 11 years. 100 of those cases were seen in the University of Heidelberg where 22 000 women received the yearly serial screening examination, that included physical, mammographic and telethermographic examination. In this group 792 breast cancers were discovered, 631 (80%) were diagnosed on the initial examination.

On the 161 remaining cancers 100 cases showed measurable tumour nuclear shadow in the foregoing mammography. An additional 40 cases showed physical or thermographic abnormalities in the previous examinations, so that together 140 cases (87%) are prevalence cancers. The remaining 21 (13%) cancers showed no evidence of their presence on previous screenings. These are the 13% 'true' incidence cancers in the Heidelberg screening group.

The volume doubling time in cases with measurable tumour nuclear shadow was obtained by measuring the growth of this tumour during the time interval of 2 screenings.

Limitations of Method

This method is subject to various sources of error, detailed criticism has been published by FOURNIER (1977) and may be summarized as follows:

1. The tumour-cell-stroma-relationship varies and therefore parts of the tumour definitely consist of stroma.

2. Very fast growing and very slow growing carcinomas could not have been included in the study.
3. The tumour nuclear shadow cannot be precisely defined, measurement of the tumour is subject to error.

But it seems sufficient that the deviations from correct measurement are consistent with each mammography.
The detailed biometrical data analysis is given by FOURNIER and WEBER (1977) and SPRATT (1978).

Results

95% of the observed volume doubling times T_v lie between 65 and 627 days. The mean of all doubling times is 212 days with a 95%-confidence limit of 191 days and 235 days.

Relation between different tumour diameters and age of patients

Regarding the doubling times it was estimated what age a woman could have had if the tumour would have had the following size:

Woman with a 20 mm in size tumour had a mean age of 58 years. The mean age of woman with 0,1 mm large tumour was 44 years and those with a first tumour cell of 0,01 mm size were 35 years old.

These extrapolations may have considerable consequences concerning clinical questions like: Time interval between 2 mammographies, age at which mammography should be started and the value of treatment in comparison to the natural life span of a tumour.

For clinicians the main results are:
1. It needs on average 20 years to reach a tumour size of 2 cm.
2. In the most cases the growth of breast cancer starts in the age interval of 30-40 years.

Relations between speed of growth, age and incidence of axillary lymph node metastases

In 125 cases satisfactory histological findings in the axillary lymph nodes could be found.
In the age group of less than 50 years (45 women) it could be seen that 14 (31%) of them have had lymph node metastases, whereas the remaining 31 cases (69%) have had negative lymph nodes.

The speed of tumour growth was significantly faster in the 14 cases with metastases (T_v = 179 days) when compared to the 31 cases without

metastases (mT_v = 226 days).

In the group of patients older than 70 years of age (11 women) all 11 cases showed free lymph nodes.

The average speed of growth was significantly lower (mT_v = 244 days) in the older group in comparison to women of less than 50 years of age (mT_v = 210 days).

12 cases with 5 or more mammographies per case

Those 12 cases observed for many years showed no significant changes in their speed of growth. The observed growth indicates very well the existence of an average exponential growth in the observed period of tumour life.

Fast, slow and moderate growth

The 13% 'true' interval cancers were mostly of the fast growing type and so they were not measurable in serial mammograms. On the other hand very slow growing tumours are also missed in frequency distribution of growth rates, because they do not show measurable changes in size for years.

In the measurable group of 147 cases we observed doubling times of less than 150 days in 28% and more than 150 days in 72%.

Growth rate and histological diagnosis

No correlations between histological diagnosis and speed of growth could be found.

Growth rate and telethermographic findings

Pathologic-thermographical signs occurred more frequently in tumours with a rapid growth rate.

The group of faster growing tumours (T_v < 150 days) was thermographically suspicious in 70%, whereas the group of slower growing tumours (T_v > 150 days) was suspicious in 41% only.

Discussion

Very fast growing breast cancers are mostly not measurable by the method used. They occur during the used screening-interval and are 'true' interval cancers (13%).

They cannot be detected 'early' by current methods such as mammography.

On the other hand very slow growing types are not measurable also, because they do not change the tumour size for years.

But in the main group of breast cancers the growth rate was measurable and showed a mean of volume doubling time of 212 days. All the cases with observation times of 0,2 - 11 years showed exponential growth between tumour sizes of 0,2 mm to 70 mm. There are no indications that the growth behaviour of those tumours are others than exponential in the earlier stages before 0,2 mm in size.

On the other hand theoretical (ARCHAMBEAU, 1970) and clinical evaluations (SPRATT, 1978, KUSAMA, 1972) showed the possibility of the 'Gompertz'-function in the very late phase of a tumour with a weight of some kilogramms.

The observed doubling times as well as age of patients and sizes of tumours showed a log-normal distribution.

These results correspond with observations of GERSHON-COHEN (1963), KUSAMA (1972) and SPRATT (1977).

It could be shown that patients with axillary lymph node metastases have significantly higher speeds of growth than those with free lymph nodes. In older patients the speed of growth is significantly slower and their incidence of lymph node metastases is significantly lower than in the group of younger patients.

It seems that the observed effective growth rate is the net result of the foregiven cell-dividing rate and of growth inhibiting factors on the other side. The importance of these factors probably will increase when the tumour becomes larger (ARCHAMBEAU (1970), BLOOM (1962), HEUSER (1978), HOEFFKEN (1977)).

Conclusions

1. About 13% of the very fast growing 'true' interval cancers are missed by the 3 screening methods used (physical, thermographic and mammographic).
2. The majority of breast cancers requires on the average 20 years to grow to a tumour of 2 cm in size. So the mammographic screening should be performed every two years.
3. Because of the fact that the majority of breast cancers may start invasive growth in the 30 - 40 years age group the screening should start in this age.
4. Therapeutic results should be judged 15 - 20 years after treatment regarding the mean volume doubling time of 212 days.

5. Thermographic pattern of the tumour may allow statements on biologic activity and growth rate.
6. According to SPRATT (1978) it seems that the natural growth rate of a tumour has much more influence on the patient's life than all other factors like early detection and course of treatment.

Literature

D.v.Fournier et al.: Wachstumsgeschwindigkeit des Mamma-Karzinoms: Konsequenzen für Frühdiagnose, Therapie und Nachsorge
in: E. Krokowski 'Neue Aspekte der Krebsbekämpfung
Georg Thieme Verlag Stuttgart (1979) 39-52

Spratt, jr. J.S., M.L. Kaltenbach, J.A. Spratt:
Cytokinetic Definition of Acute and Chronic Breast Cancer
Cancer Research 37 (1977) 226

Hoeffken, W.,K. Heuss: Beobachtungen zur Wachstumsgeschwindigkeit beim Mammakarzinom.
Vortrag: Deutscher Röntgenkongreß, Münster, 19. - 21. Mai 1977

Archambeau, J.O., M.B. Heller, A. Akanuma, D. Lubell:
Biological and clinical implications obtained from the analysis of cancer growth curves.
In: Knapp, R.C. ' New concepts in gynecologic oncology' 13 (1970) 831

Comments to Dr. v. Fournier's paper

1

Dr. Roberts

I agree that screening for breast cancer would be better done at an earlier age. The problem is that mammography is undesirable in women under the age of 35 years, especially if done repeatedly. There is much evidence showing that the breast tissue of young women is highly radiosensitive and the actual incidence of breast cancer could be increased by repeated exposure to X rays. There is probably no such thing as a threshold dose.

Answer to Dr. Roberts

2

Dr. v. Fournier

The conventional film-mammography is still the only method able to reproduce the finest calcifications and requires an exposure of 3 rads integral dose. The risk of carcinogenesis due to such an exposure is unknown but should be avoided.

The new technique of 'screen-mammography' produces an acceptable quality of result with a dose of only 0.1 to 0.3 rads. The way to give safe and efficient reduction of breast cancer mortality in the younger age groups is by a reduction of the dose needed in mammography rather than by renouncing screening completely.

3

Dr. Haybittle

You have found, as others reported, a lognormal distribution of doubling times. It is perhaps interesting that if one takes a very simplistic model of the effect of unsuccessful treatment, namely that a tumour is reduced to a certain residual size, then grows again with its original growth rate until at some arbitrary size the patient dies, a lognormal distribution of doubling times will lead to a lognormal distribution of survival times as in fact occurs in many series and was first pointed out by Boag in 1948.

Answer to Dr. Haybittle

4

Dr. v. Fournier

The observed lognormal distribution of doubling times comes from measurements on clinically 'relevant' cancers only and this may not directly relate to survival

5

Dr. Ribeiro

I would like to raise the question of what we should do with patients who have a particular high risk of breast cancer. I feel that this is a very difficult area.

Answer to Dr. Ribeiro

6

Dr. v. Fournier

The risk may range from a factor of 4 in cases with mastopathia including proliferations and atypic cells to a factor of about 40 as in a case where the mother and two sisters show premenopausal breast cancer. So in relation to the risk in cases with a highly increased risk factor (breast cancer in two premenopausal sisters) a definate solution by preventative subcutaneous mastectomy (with subsequent subpectoral augmentation) may be recommended. In cases with increased risk and significant cancerophobia this operation can be performed also.

Ablatio mammae with subpectoral augmentation and free transplantation of the areola may be an alternative method which removes the breast tissue safely.
About 5% of women with high risk factors wanted such a definate solution of the problem. The remaining group with a lower risk factor underwent subsequent mammographical, physical and thermographical controls at time intervals of 1 to 3 years.

7

Dr. Baum

Whilst enjoying Dr. v. Fournier's paper very much I feel that there are a number of false assumptions in the development of his argument. Firstly, he is only measuring those tumours which are well defined (i.e. with pushing margins). These must, therefore, be a selected population as it has been suggested already that tumours with clearly defined margins behave differently from those with ill-defined margins. Secondly, there is an assumption that the measured tumour mass consists of a sphere of cancer cells. This is in fact not the case as all such tumours are made up of vascular tissue, acellular stroma and necrotic debris, in addition to the viable cancer cells.

Finally, it has been demonstrated in the past that the phase of tumour development that is measurable clinically probably does not grow according to an exponential function. However, accepting these assumptions, then the argument would suggest that screening for early breast cancer is futile, and that tumours on average would have been present for twenty years before coming radiologically detectable, and thus if they have any metastasizing potential at all, will have already disseminated.

Answer to Dr. Baum

8

Dr. v. Fournier

These points and other initial points of the method used are fully discussed by Fournier (1977)

Ref.

Fournier (1977) Wachstumsgeschwindigkeit des Mammacarcinoms und Konsequenzen für die Früherkennung und Nachsorge, Habilitationsschrift, Dekanat Med. Fakultät I, Universität Heidelberg.

The reporting of non-significant results in clinical trials

J.L. Haybittle
Physics Department
Addenbrooke's Hospital,
Cambridge.

A large number of clinical trials that have been run to date, particularly in the cancer field, have ended with a 'non-significant' result i.e. the authors have reported 'no significant difference at the 5 per cent level' or some equivalent statement. Sometimes this conclusion may not even be accompanied by a P-value, so that results with P=0.06 are classed in the same category as those with P=0.40, a most unhelpful state of affairs from the point of view of the reader who is trying to plan a future clinical strategy taking into account the trial results. I think most of us would accept as essential the statement of a P-value since the value of P=0.05 for division between 'significant' and 'non-significant' results is to a large extent arbitrarily chosen and should not be considered as a hard and fast yes/no decision point on the comparability of the effectiveness of two treatments.

But I would submit that even more information should be given routinely with 'non-significant' results so that it is made quite clear that 'no statistically significant difference' does not mean that no real difference exists. This may be done in the first place by quoting the 95% confidence limits within which the magnitude of a possible difference may still exist (Wulff, 1973). If a T-year survival rate is estimated in each group by a lifetable analysis, then these limits can be found by the standard method of estimating a standard error of the difference in the two rates. The standard errors, $(S.E.)_1$ and $(S.E.)_2$, of each survival rate, P_1 and P_2 can be calculated in the lifetable analysis by Greenwood's formula, provided that the time of interest is not towards the end of a lifetable curve where a long flat region exists (due to few patients as risk and no deaths), or by the conservative approximation given by Peto et al., (1977). The standard error of the difference of two rates, $(S.E.)_{Diff}$, is then given by:-

$$(S.E.)_{Diff} = \sqrt{(S.E.)_1^2 + (S.E.)_2^2}$$

and the 95% confidence limits of the difference given by:-

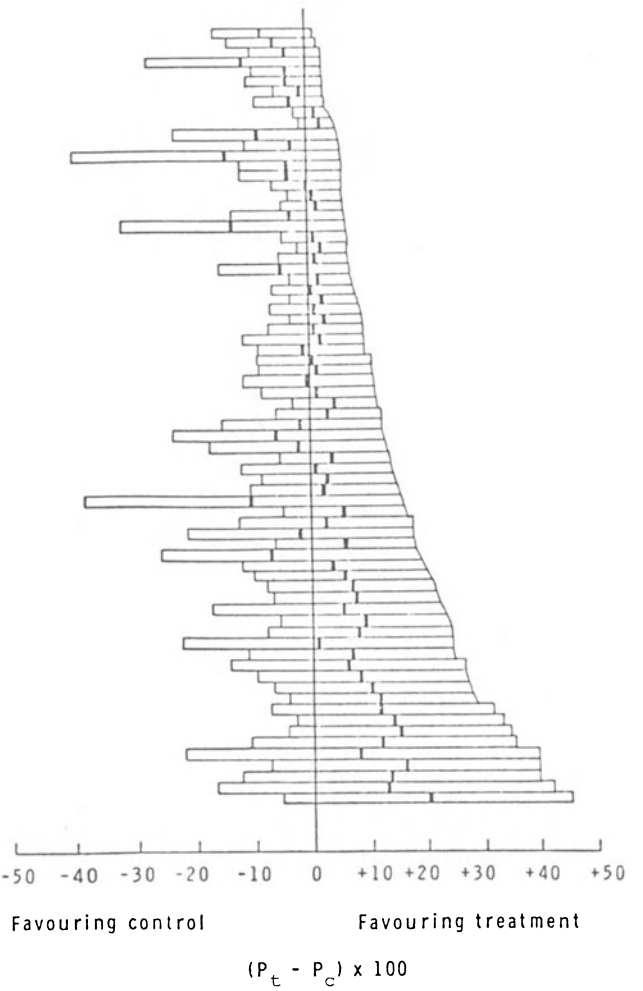
$$(P_1-P_2) \pm 1.96 \times (S.E.)_{Diff}$$

Recently Freiman et al.(1978) have made this kind of calculation, but using 90% confidence limits, for 71 clinical trials which had been reported with non-significant results. The 71 were drawn from 300 trials published in the literature during the past 10 years, in 110 of which the authors had stated that there was no significant difference or some equivalent statement. The 71 analysed by Freiman et al., were those in which the difference was not even significant at the 10 per cent level of probability on a two-tailed test. Fig. 1 shows their results. The horizontal bars are the 90% confidence limits of the differences, the actual value of the difference found being shown by the central mark in each bar. The first thing that this diagram demonstrates is the large real differences that might still have existed in all of these trials in spite of the authors reporting no significant difference. If the authors had included a statement of these limits in their results, then the reader would have been given a much clearer impression of how well the alternative to the nullhypothesis had been tested (and would in many cases have probably decided that the test had been inadequate). The large span of the 90% confidence limits is, of course, due to the small numbers of patients admitted to most of these trials, and this is the point which Freiman et al. are most concerned to bring out in their paper. Only 4 of the trials concerned stood a chance of 90 per cent or more of detecting a 25% difference in control rates.

A second point that is brought out by this presentation of confidence limits arises where there are reasons other than treatment effectiveness for preferring one treatment to another e.g. because it is a less radical procedure and therefore less disturbing to the patient. Suppose the new treatment in the lowest trial of the diagram was such a conservative procedure, and the control a more radical one. The trial shows a 20% improvement for the conservative procedure but the difference is non-significant, P=0.189. A clinician may well ask:'If I change over to the conservative procedure, what risk do I take of my results becoming worse?' An approximate answer to this question, provided that the numbers in the trial are reasonably large, is given by half the p-value quoted above, i.e. about 0.095. Also, looking at the confidence limits he can see that the chance of his results becoming more than 5% worse is 1 in 20, i.e. half the chance, 1 in 10, of lying

Figure 1

90 per Cent confidence limits for the true percentage difference of 71 trials.*



*Reproduced by permission from Freiman et al., 1978, New England Journal of Medicine, 299, 690-694

outside the limits in either direction. On the basis of these two observations he might well decide that the conservative procedure should be adopted.

But suppose the result had been as in the uppermost trial of the diagram where the difference is also non-significant with P=0.144. The 90% confidence interval extends to a difference of 17% in favour of the radical procedure, and an approximate answer for the chance of the conservative treatment giving worse results is 1 - 0.144/2 (because the observed difference is in favour of the radical treatment). On the basis of this high chance (P=0.928), the clinician might be very unprepared to adopt the conservative treatment. Thus two results, both 'non-significant', can give very different guides to future action.

The comparison of survival rates at a particular point of the life-table is not the most efficient method for testing the difference between two treatments in clinical trials requiring prolonged observation of each patient, and the logrank or an equivalent test which makes a comparison of survival experience throughout the period of observation is usually preferred. We can estimate our confidence limits from such an analysis using the following procedure suggested by Richard Peto (1978). Suppose $\lambda_1(t)$ and $\lambda_2(t)$ are the hazard rates at time t and that θ is the hazard-ratio of group 1 to group 2 throughout the period of observation i.e.

$$\lambda_1(t) = \theta\, \lambda_2(t)$$

This implies that the survival probabilities $P_1(t)$ and $P_2(t)$ are related by the equation $P_1(t) = P_2(t)^{\theta}$.

If θ is no very different from unity Cox (1972) has shown that the maximum likelihood estimate of θ is given by:-

$$\log_e \hat{\theta} = (O_1 - E_1)/V_1$$

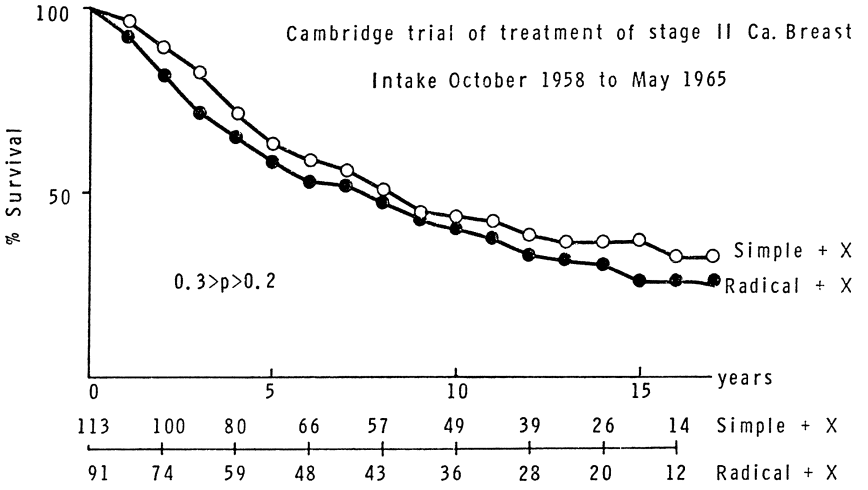
where O_1, E_1 and V_1 are the observed and expected deaths in group 1 and the variance of (O_1-E_1) respectively. The 95% confidence limits for $\log_e\theta$ are given by:-

$$\frac{O_1-E_1}{V_1} - \frac{1.96}{\sqrt{V_1}} \quad \text{and} \quad \frac{O_1-E_1}{V_1} + \frac{1.96}{\sqrt{V_1}}$$

Let us look at the application of this to a clinical trial of the treatment of breast cancer which produced a non-significant result. Fig. 2 shows the results of the Cambridge trial of Simple versus Radical Mastectomy in Stage II cancer of the breast. Both groups had

Figure 2

Survival curves of Cambridge Trial of Radical mastectomy v. Modified Simple mastectomy, both followed by X-ray therapy, in the treatment of Stage II cancer of the female breast.



post-operative X-ray therapy. The result of the logrank test on this data was $\chi^2 = 1.153$; $P = 0.28$. Taking the radical group as group 1 (where the hazard rate was higher i.e. the survival rates were a little lower throughout the period),

$$O_1 = 67,\ E_1 = 60.685 \text{ and } V_1 = 34.52$$

This leads to $\log_e\hat{\theta} = 0.183$, $\hat{\theta} = 1.20$ with 95% confidence limits for $\log_e\theta$ of -0.150 and 0.516, corresponding to limits of θ of 0.86 and 1.68. Unfortunately, clinicians are not used to dealing with hazard ratios so these figures may not be very meaningful to them. It may surprise them for instance that an increased hazard rate of 20% (i.e. a hazard ratio of 1.20) only gives rise to a reduction of about 5% in a 5 year survival rate of about 60%.

We can however use the confidence limits for θ to obtain the confidence limits for 5 year survival rate in the radical group assuming the rate in the simple group to be the observed value. The predicted 95% confidence limits for the 5 year rate in the radical group are $100 \times (0.6283)^{1.68} = 45.9\%$ and $100 \times (0.6283)^{0.86} = 67.1\%$. These are shown in fig. 3 together with the confidence limits predicted by the standard error method described earlier. The results of calculations for the 10-year survival rates are also given. The two methods give similar but not identical results, the calculations from θ resulting in a smaller range between the limits, reflecting the greater efficiency of the logrank comparison. We can thus add some useful information to guide the clinician's future policy. He has a probable range of survival rates from radical treatment compared with that from simple, and can see that there is only a 1 in 40 chance of radical improving the 5-year and 10-year rates by more than about 5%. Also the probability of simple treatment being at all worse than radical is approximately 0.14 i.e. half the P-value found in the comparison.

In this trial the confidence limits were quite large because of the small numbers of patients involved (204). By contrast the CRC Breast Trial comparing simple mastectomy + post-operative DXT with simple mastectomy + watch policy has very large numbers entered for it. Fig.4 shows the survival results at October 1976. The difference was decidedly 'non-significant', $\chi^2 = 0.0027$; the hazard ratio $\theta = 0.994$. The 95% confidence limits of the watch policy 5-year rate calculated by the two methods already described are shown in fig. 3. The limits calculated from θ are again smaller than those calculated from the standard errors, and show that it is very unlikely (less than a 1 in 20 chance) that any real difference between the treatments would

Figure 3

95 per cent confidence limits of survival rate in group having more radical treatment predicted from the rate in the other group by calculations based on either the ratio, θ, of the hazard rates, or on the standard error of the difference of the two survival rates.

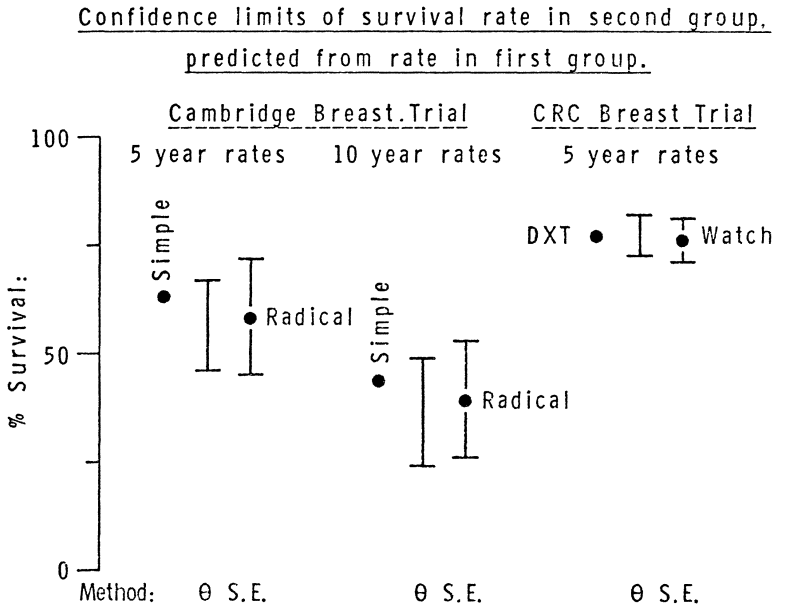
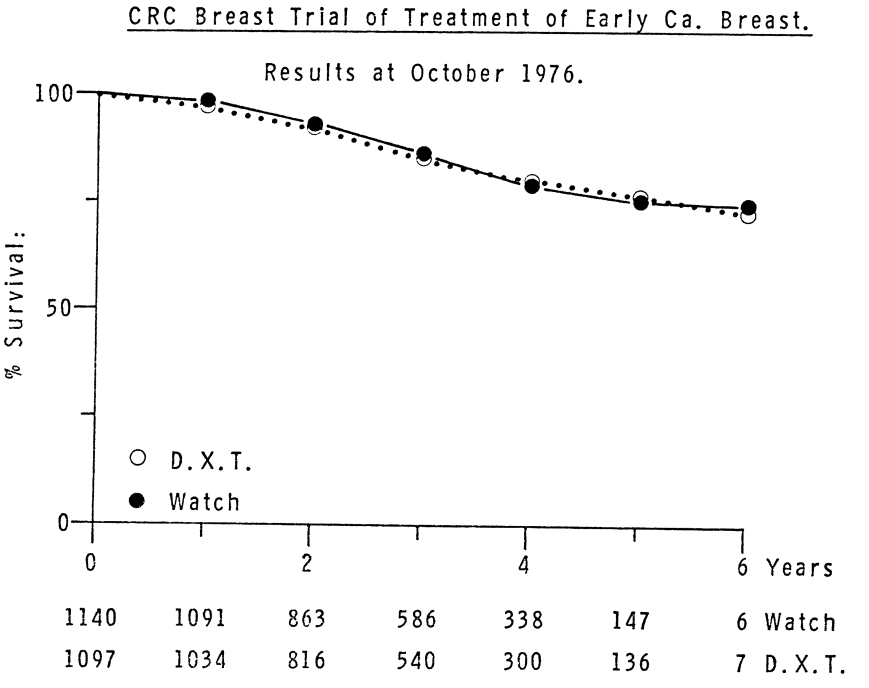


Figure 4

Survival curves of the Cancer Research Campaign (CRC) Trial of Simple Mastectomy + X-ray therapy (DXT) v. Simple Mastectomy alone (Watch) in the treatment of early cancer of the female breast. Curves based on data available in October 1976.



result in an absolute difference of 5 year survival rate of more than ± 4.5 per cent. The large numbers have narrowed down the range of uncertainty as compared with that in the Cambridge trial. Note however that if the clinician asked the question: 'What is the chance of Watch Policy (the more conservative treatment in this case) being worse than DXT?' He would be told about 1 in 2, much higher than that given in answer to the similar question posed in the Cambridge trial, because the P-value resulting from the comparison of the two curves is so high (=O.96).

Fig. 5 compares the answers the two trials would give concerning the probability that adoption of the more conservative treatment will lead to different decreases of 5-year survival rate. In this case the decrease, X, is expressed as a percentage of the observed rates found in the trials. Both trials suggest that there is an approximately 1 in 2O chance of the adoption of the more conservative procedure leading to a decrease of X = 5% in 5-year rate. Of course, in deciding future policy from both trials the clinician has other information to take into account e.g. in the case of the CRC trial the incidence of local recurrence was significantly reduced by DXT.

These are just two examples of the way in which 'non-significant' results can be more fully documented, and made in my view more helpful to the clinician in making judgements about future policy. In conclusion I would suggest that the presentation of a 'non-significant' result of a clinical trial, should include:

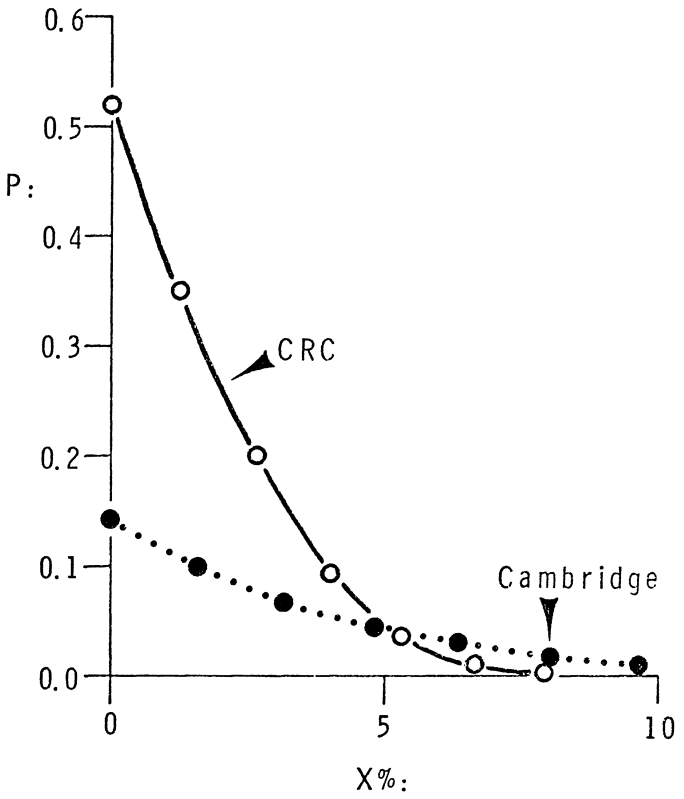
(a) a statement of the exact P-value found in a logrank or similar comparison,
(b) the 95 per cent confidence limits deduced by a hazard ratio calculation for a survival (or recurrence) rate of one group in comparison with the rate in the other group,
(c) if one treatment is less desirable than the other for reasons other than its effectiveness, an indication of the probability of this treatment being better than the other.

Acknowledgements

I am greatly indebted to Mr. Laurence Freedman of the M.R.C. Clinical Trials Office, Cambridge, for most helpful discussions on the subject of this paper.

Figure 5

Probability of more conservative treatment resulting in an X% decrease in 5 year survival rate, where X is expressed as a percentage of the observed rates found in the two trials.



P = probability of more conservative treatment resulting in an X% decrease in 5 year rate.

References

Cox, D.R., (1972) Regression models and lifetables.
Journal of Royal Statistical Society, B., 34, 187-202

Freiman, J.A., Chalmers, T.C., Smith, H., and Kuebler, R.R. (1978)
The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial.
New England Journal of Medicine, 299, 690-694

Peto, R. (1978) Private communication.

Wulff, H.R. (1973) Confidence limits in evaluating controlled therapeutic trials.
Lancet, ii, 969-970

Comments to Dr. Haybittle's paper

1

Dr. Crowley

A technical point or two:

1) The confidence interval reported for the survival curve using proportional hazards is likely to be too small, as it takes account of only the variation in the ratio of hazards estimate and not the survival curve.

2) The confidence interval for the ratio of hazards is also approximate, as it is based on ignoring certain terms in the variance (which are quite small near the null hypothesis).

Answer to Dr. Crowley

2

Dr. Haybittle

I agree with you and I was very careful in my talk to use the word 'approximate' when estimating the confidence limits. But if the hazard ratio is not too different from unity then I think the approximations are good enough for the purpose of trying to give the clinician more help in interpreting the result of a trial.

3

Dr. Prentice

A minor technical point is that confidence intervals for 5-year survival, for example, may be more accurate if such confidence intervals were formed not by attaching standard errors to $\hat{p}$, the estimated 5 year survival rate, but rather to some transformation of $\hat{p}$, such as $\log(-\log \hat{p})$, that would be free of range restrictions.

Notes on Clinical Trial Methodology

Richard Peto
Reader in Cancer Studies, Oxford University

References

The chief points that I wish to make about clinical trial methodology have already been published, either in our joint article on clinical trial methodology (British Journal of Cancer 1976, 1977: Design and Analysis of Randomised Trials Requiring Prolonged Observation of Each Patient. Part I:34, 585-611 and Part II:35, 1-39) or in a more recent article on this same subject (Biomedicine 1978: Clinical Trial Methodology, 28, 24-36). Because the Biomedicine article discusses in detail the main aspects of trial design which I wish to touch on today, I have circulated reprints of it to all participants at this meeting. Further aspects of trial design, and an account written for nonstatisticians of just those few statistical methods which I feel that they should be really familiar with, may be found in the Br.J.Ca. article.

Trial Design

It seems to me that the faults of clinical trial design which most impede medical progress are currently:

(1) Making a comparison which is not scientifically very interesting. Statisticians can only help here by insisting on being told enough about the disease for it to become clear why a particular trial is necessary, as in the process of explaining this the investigators may change their aims usefully.

(2) Undertaking a trial of inadequate size, so that purely random differences between the patients in two different treatment groups could well swamp medically significant real differences between the efficacies of the two treatments. Most British cancer trials, even in common cancers such as breast, have fewer than 100 patients in them, which is grossly inadequate, and I suspect that the situation is similar elsewhere. Methods of recruiting larger numbers are discussed in Appendix 1 of the Biomedicine paper.

(3) In circumstances where proper randomisation would have been practical, choosing unrandomised designs or excluding "protocol deviants", etc. from the final analysis. Both these practices can cause serious bias, sufficient for a real treatment benefit to be unne-

cessarily missed or for a spurious difference between two treatments to be generated. Reasons for preferring randomised designs have been reviewed by many authors, and there is a real example in the Biomedicine paper where the repeated use of nonrandomised designs helped to delay the proper evaluation of a potentially important therapeutic strategy for a quarter of a century.

Reasons are given in Section 13 of the Br.J.Ca. paper for requiring that any final publication of a trial should contain (among other analyses, perhaps) at least one report of the overall outcome in all patients allocated to each treatment group, irrespective of whether or not they actually received that treatment.

(4) Finally, a less fundamental point but one whose implementation is unusual in that it yields obvious benefits for very little extra effort, people should be far more ready to consider "2 x 2" or other "factorial" trial designs. These are discussed in Appendix 2 of the Biomedicine paper, and usually offer two answers for the price of doing only one clinical trial. They would be especially relevant to the investigation of hormone and chemo-therapy in breast cancer.

Trial Analysis

Statisticians will use many different methods to investigate the data generated by clinical trials, but if at all possible they should present their findings to the non-statisticians who have to actually use the trial findings in extremely simple ways. The techniques which best combine simplicity with statistical efficiency are:

(a) Life-tables (synonymously also called "Kaplan-Meier", "actuarial", "product-limit", "experimental survival" or just plain "survival" curves), to describe differences between groups of patients. Physicians engaged in clinical trials should be familiar with the interpretation of survival curves, as described, for example, in Section 18 of the Br.J.Ca. paper.

(b) Logrank tests (synonymously also called Mantel, Mantel-Haenszel, Mantel-Peto-Cox or even Savage-Mantel-Peto-Cox) to derive P-values which test for statistical significance any difference between groups of patients. The O's and E's derived in computing a logrank test are also moderately useful descriptive aids, of course, and physicians engaged in trials should be familiar with their format and interpretation as described, for example, in Sections 19-20 of the Br.J.Ca. paper.

(c) Retrospective stratification, cautious subdivision of time into "early" and "late", and cautious examination of particular subgroups of patients are also useful extensions of (b), and are described in Section 22 and 25 of the Br.J.Ca. paper.

Apart from familiarity with life-tables, logrank P-values, retrospective stratification (and its relevance to 2 x 2 designs), and the real need for randomised trials to be large, I do not feel that non-statisticians engaged in trials need familiarise themselves with any statistical methods, and indeed trying to do so may produce an unnecessary feeling of confusion. With the unfair benefit of writing these notes a few days after presenting my talk and hearing the discussion of it, I want to emphasise that I am not trying to limit what other statisticians do, and that I do consider Cox's methods (with the same need for time subdivision that exists in logrank methods) optimal in many senses. However, the identity of the logrank (O-E) and its variance with certain derivatives of Cox's likelihood function (see Statistical note 7 in the Br.J.Ca. paper) means that many of Cox's virtues can be acquired more simply by logrank observed and expected numbers rather than by likelihood manipulations, and I would encourage this simplicity in that large majority of cases where it does not conflict with the qualitative scientific understanding of the data.

Comments to Dr. Peto's paper

1

Dr. Kalbfleisch

Dr. Peto has repeatedly expressed his opinion that the life table method and logrank test should, to the exclusion of all other approaches, form the basis of the analysis of failure time data.

The statistician involved in a clinical trial analysis is confronted with two basic problems. First he must acquaint himself with the problem and the data and, using various exploratory techniques, identify and quantify to himself the important features. When this has been done, he must communicate, in a clear and unambiguous manner, the main result of this analysis to the clinician. This process should continue as the statistician and the clinician jointly develop the analysis.

The logrank method was first derived in 1966 by Mantel using entirely intuitive arguments; it is a simple test and is highly effective for the comparison of treatments with or without some stratification. The intuitive level on which the test can be presented makes it an ideal

tool for communication. If Dr. Peto's recommendations where to use the logrank test for this purpose, I would support them. Dr. Peto's comments today, however, seemed to suggest that statisticians should use only the logrank test and none of the other methods of analysing failure time data. The Cox model, from which the logrank test can be derived, is however a much more powerful tool. Its use allows the statistician to look at several covariates simultaneously and adjust failure rates for them. In addition, it is simple to carry out a routine scanning of the data for possible interactions with treatment and is in general a more convenient and efficient approach to data analysis than is the logrank procedure. Regression models have served many areas of statistical application well and should not be discarded in clinical trials. The statistician needs many advanced tools to carry out effective data analysis. To confine our attention to only two - the Kaplan-Meier approach to life table estimates and the logrank test - would be foolish indeed.

2

Dr. Prentice

It is essential in a discussion of statistical methods for the analysis of clinical data to distinguish between methods used to analyze and explore a data set and methods used to present the findings of such analysis. The material presented, concerning statistical techniques that should or should not be utilized, does not adequately distinguish these two topics.

3

Dr. Roberts

(1) Stratification is best done on entry to a trial, rather than retrospectively, because it is more likely that essential data will be recorded accurately by the clinician if it is required for the randomisation procedure.

(2) There could be a biological difference between tumours in premenopausal and postmenopausal women which might be an important reason for differences in results of chemotherapy. We are neither consistent nor accurate in recording menstrual status, so I wonder if the Milan chemotherapy data have been analysed according to age decades, to determine if the differences are only between very young and definitely postmenopausal and whether those nearer the menopause in either direction are different.

4

Dr. Haybittle

I am very concerned at the interpretation of the P-value obtained in a subgroup when the logrank test is used on the data subdivided according to some prognostic factor. We had an example yesterday from Dr. Wallgren in the results of the Stockholm trial, where the overall comparison was 'non-significant', but in the group with medial tumours there appeared to be some advantage for preoperative radiotherapy, and when a comparison was made in this group alone the P-value was just under 0.05. Shouldn't we be looking for much lower P-values in a subgroup, if the overall comparison is non-significant, before placing too much weight on such a result?

The Cardiff Mastectomy Trial

by
M. Maureen Roberts
Dept. of Clinical Surgery,
University of Edingurgh

Introduction

The Cardiff Mastectomy Trial was initiated ten years ago by Professor A.P.M. Forrest, with the objective of comparing a conservative policy of treatment based on axillary node histology with a standard radical policy of treatment in patients with primary breast cancer.(1,2) Although the surgical policy under test was a simple mastectomy, node histology was considered essential as clinical staging of the axilla is known to be inaccurate. (3)

Definition of Treatment Policies

In the conservative policy, simple mastectomy was performed, removing all of the breast and its axillary tail together with a biopsy of the lower-most pectoral nodes which are in continuity with the nodes in the axillary tail. (4) If any node was infiltrated with tumour, then postoperatively radiotherapy was given to the axilla alone, consisting of 4,000 rad from a cobalt source in 10 fractions given on alternate days over a three week period. If there was no evidence of tumour spread to the pectoral nodes, then nothing further was advised, the patient being treated by simple mastectomy alone.

In the radical policy of treatment, the surgeon was allowed to perform a radical mastectomy of any type, providing that full axillary clearance was achieved. If the nodes in the axilla were involved with tumour then radical post-operative radiotherapy was given, consisting of 4,000 rad to the chest wall, 3,500 rad to supraclavicular and internal mammary regions, and 4,000 rad to the axilla given in 10 fractions, carried out over a period of four weeks. If nodes were free of tumour, then radiotherapy was not given.

Each of the two policies of treatment therefore had two sub-groups depending on the histology of the axillary nodes.

Structure of the Trial

Patients with primary breast cancer of international stage 1 and 2 (T1 or T2, NO or N1, MO) were admitted to the trial, and were then

stratified, and randomly allocated to either of the two policies. Only patients with stage 3 or 4 disease or with tumours arising during pregnancy or lactation were excluded from the trial and there was no upper age limit. Stratification was carried out according to the clinical stage of the tumour, palpability of nodes, site of tumour in the breast (medial or lateral), and menstrual status. Menstrual status was defined as pre-menopausal (periods regular, or up to two years since the last menstrual period), menopausal (two to five years since the last menstrual period), and post-menopausal (over five years since the last menstrual period). Patients who had undergone a previous hysterectomy were included in one of the menstrual groups according to the time interval since the hysterectomy was performed.

Results

From October 1967 until June 1973 a total of 200 patients were admitted to the trial, with 103 allocated to the conservative and 97 to the radical policy. Two pairs of treatment groups were available for comparison: 64 patients who had simple mastectomy alone and 66 patients who had radical mastectomy alone, 39 patients who had simple mastectomy and axillary node radiotherapy and 31 patients who had radical mastectomy and radical radiotherapy.

Comparison of Groups

Patients were well matched according to the four clinical variables of stratification, and for age, with a total range within the trial of 30 to 88 years. It is worth noting that only 21 patients in the trial (10%) had tumours of size 2 cm or less; only 30% patients in the trial were pre-menopausal.

Analysis

The results of the trial have been analysed according to survival, recurrence patterns, and morbidity of treatment.

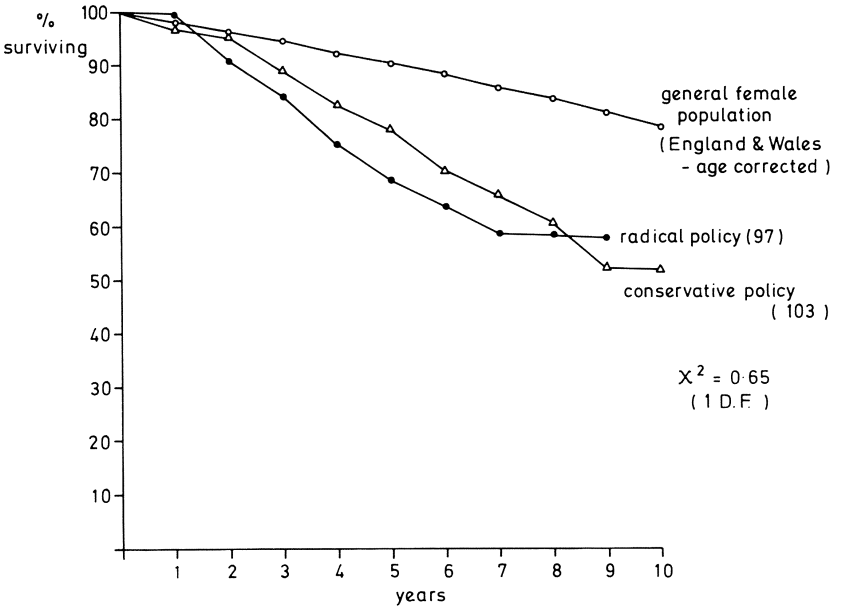
The current status of patients in the trial (April 1978) is shown in Table I. A total of 94 (47%) are alive and free of recurrence with a minimum follow-up of 5 years and maximum of 11; only 7.5% of patients have died of causes other than breast cancer, without any evidence of recurrent disease. The percentage of patients surviving each year following treatment is shown in Figure 1, compared with the survival of the general female population of England and Wales, derived from the Registrar General figures and age corrected to match patients

Status of Patients	Total	Policy of treatment	
		Conservative	Radical
alive, no recurrence	94	46	48
alive, with recurrence	31	21	10
dead, with recurrence	60	29	31
dead, no recurrence	15	7	8
Total	200	103	97

TABLE 1

The current status of patients within the trial, according to policy of treatment. The trial started in October 1967, and the last entry was in June 1974; figures analysed April 30th 1978

Figure 1 Survival curves of radical policy and conservative policy groups of patients, analysed by log rank test using computer programme due to Pike (by courtesy of Medical Computing and Statistics Unit, Edinburgh of Edinburgh).



within the trial. There is no significant difference between patients treated by the conservative policy and those treated by the radical policy, when analysed by the Log Rank test. (5) The survival of patients in the two sets of treatment sub-groups are shown in Figures 2 and 3. Patients who were treated by surgery alone fared equally well whether radical or simple mastectomy was performed. Although there appears to be a slight advantage for conservative treatment in patients who had positive nodes, (Figure 4) log rank analysis shows no significant difference between the treatment groups, and at ten years the survival curves are identical with only 40% of patients in either group still alive.

Recurrence

The site of first recurrence (April 1978) has been analysed for each of the four treatment groups, Table II. We were particularly interested in local recurrence, and found that patients treated by simple mastectomy alone had a 15% incidence of axillary recurrence, whereas none of the patients treated by radical mastectomy presented with axillary recurrence as a first event. The incidence of dissemination was identical in the two groups.

In patients treated by surgery and radiotherapy, a significantly increased scar recurrence was found in those patients treated by mastectomy with axillary radiotherapy, compared with patients treated by radical surgery and radical radiotherapy which included the chest wall. Disseminated disease the first indication of recurrence occurred slightly more often in patients treated by radical surgery and radical radiotherapy.

Morbidity

As part of the trial an assessment of morbidity was made in a sample of patients from each of the treatment groups. This included objective measurements of shoulder movement and arm oedema. This has been reported elsewhere (2) but an example of the results is shown in Table III. The incidence of both arm oedema and restricted shoulder movement increased with the extent of treatment.

External review

We have analysed the results of the trial in terms of survival, site of first recurrence and morbidity, according to the treatment which the patient received. In April 1978, because it was the tenth year

Figure 2 Survival curves for patients treated by surgery alone, analysed as for Figure 1.

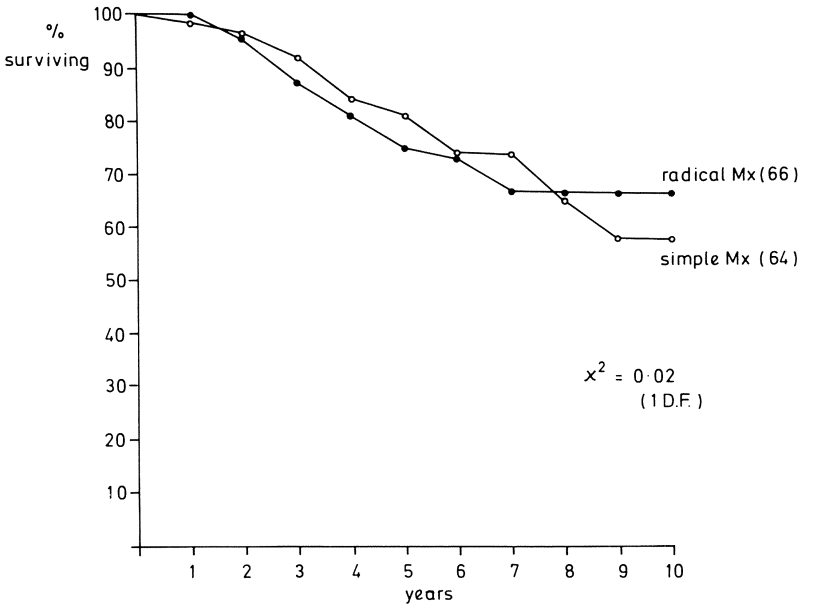
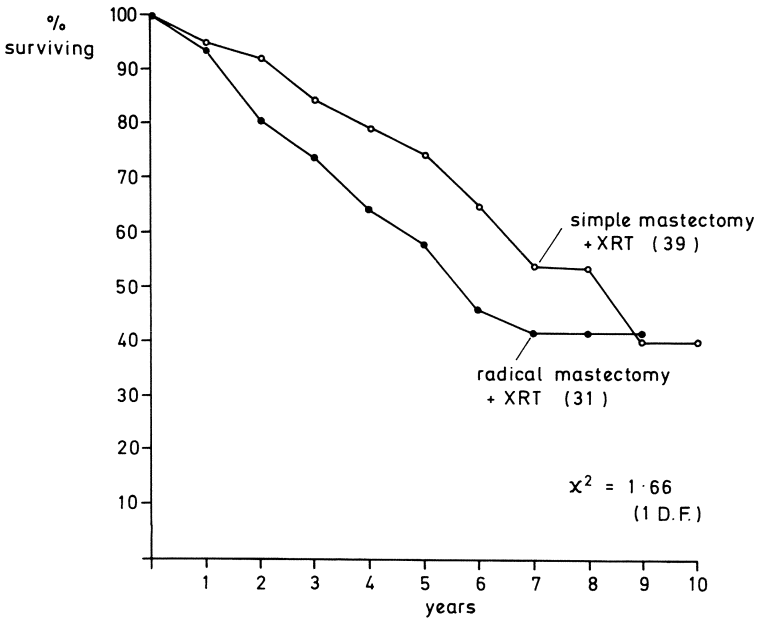


Figure 3 Survival curves for patients treated by surgery and radiotherapy, analysed as for Figure 1.



Type of Recurrence	Surgery Alone		Surgery and Radiotherapy	
	Simple Mx (64)	Radical Mx (66)	Simple Mx + XRT (39)	Radical Mx + XRT (31)
Regional				
Axilla	10*	0	0	1
Scar	6	9	10**	1
Total (Patients)	18	12	12	3
Disseminated				
Total	12	12	10	16***
Total Patients with Recurrence	29 (45%)	23 (35%)	21 (54%)	18 (58%)

TABLE II

Pattern of recurrence, according to FIRST noted site, in each of the 4 treatment groups. Analysis at April 30th 1978. Some patients had more than one site involved simultaneously. Patients treated incorrectly according to protocol remain in their designated treatment groups.

* Comparison of simple Mx and Radical Mx: $X^2 = 10.83$ $p < 0.001$

** Comparison of simple Mx with XRT to axilla and radical Mx with radical XRT: $\chi^2 = 6.97$ $p < 0.01$

*** $\chi^2 = 3.95$ $p < 0.05$

	Surgery alone		Surgery and Radiotherapy	
	Simple (28)	Radical (28)	Simple + XRT (19)	Radical + XRT (12)
arm swelling > 2 cm	* 4 (15%)	13 (46%)	7 (37%)	7 (58%)
limited elevation of shoulder	** 0 (0%)	7 (25%)	7 (37%)	8 (67%)

TABLE III

The incidence of arm swelling and restricted shoulder movement in the four treatment groups.

* Comparison of simple and radical mastectomy

$\chi^2 = 7.47 \quad p < 0.01$

** $\chi^2 = 6.02 \quad p < 0.03$

since the trial began, we decided to review all the trial records and spent a week in Cardiff with Dr. Helen Stewart acting as our external reviewer. Three further aspects of the trial will be considered, in the light of our findings. These are protocol violations, failure of local control of the disease, and disagreement over definitions.

Protocol Violations

Because our entry criteria were minimal, there were no violations on this score (Table IV). On surgical technique, 10 patients in the radical mastectomy group were thought to have had doubtful clearance of the axilla and there was failure to identify a pectoral node in 28 patients in the simple mastectomy group. There were few errors in the radiotherapeutic techniques, only 8 patients not receiving the correct treatment. Two patients who had negative nodes received radiotherapy to the axilla, 5 patients who had positive nodes were not given radiotherapy in error, and 1 patient received too low a dose. We have previously reported three of these errors, but were not aware of the others until our review.

The low number of errors with regard to radiotherapy is almost certainly because only one radiotherapist was involved, whereas a total of 37 surgeons were involved in performing the operations.

We did not know when setting up the trial whether failure to identify a pectoral node was important, merely grouping together all patients in whom there was "no evidence" of tumour spread to the axilla. Some justification for this approach is given by the figures in Table V, which shows the overall incidence of positive nodes was identical in the simple and radical mastectomy groups. It is noteworthy that the majority of the mastectomies (63) were performed by trainee surgeons, who failed to find a node in one third of cases; in the remainder, which was done by consultant surgeons, failure to find a node occurred in roughly 17% of cases.

Failure of Local Control of the Disease

From our 10 year review, we have tried to determine the <u>total</u> incidence of local recurrence, both as a first event and if it occurred subsequently. The results we found must represent a minimal incidence, for there may well have been cases in which local recurrence was not fully documented, particularly if the patient was dying of widespread disease. We have established (Figures 1,2,3 and Table II) that survival and dissemination are similar for the two treatment policies, and must

	Conservative Policy	Radical Policy
Entry Criteria	0	0
Surgical Technique:		
"doubtful" clearance	-	10
Failure to Identify Node	28	2
Radiotherapy Technique:		
Negative Node Given XRT	2	0
Positive Node no XRT	1	4
Positive Node Low Dose	1	0

TABLE IV

Protocol violations or failures in the Cardiff Mastectomy Trial.

	simple mastectomy	radical mastectomy
Total	103	97
nodes identified	75	95
positive	38	35
negative	37	60

TABLE V

The incidence of identified nodes in the two groups, and the number of these which proved to be involved by tumour.

now consider the problem of local control of the disease. Failure of local control is shown in Table VI for patients in whom there was no evidence of spread to the axilla. Over the 10 year period of follow-up, the overall incidence of local recurrence was similar in the simple mastectomy (27%) and radical mastectomy groups (23%). Although axillary node recurrence was increased overall in the simple mastectomy group, it was not related to the identification of the pectoral node at mastectomy. It is noteworthy that the incidence of scar recurrence was high (15% overall) which may be due to the high proportion or T2 tumours in the trial.

In patients in whom there was axillary involvement by tumour (Table VII) there was a significant difference between the treatment groups. In those treated by the conservative policy (i.e. radiotherapy to the axilla alone) there was a 30% incidence of scar recurrence, compared with 3% in those patients treated by the radical policy. Axillary recurrence was minimal in both groups, presumably because both policies included treatment of the axilla.

The local recurrence rate was similar whether tumours were in the lateral or medial half of the breast in both treatment policy groups.

Disagreement on Review

A critical review of our entry criteria showed that we might have excluded 26 patients on various grounds if we had defined our entry criteria more strictly. Two patients had non-invasive carcinoma of the breast, one patient had bilateral breast cancer and three others had a previous history of breast cancer, one had another malignancy previously, seven were over 75 years of age and 15 had undergone hysterectomy, some with oophorectomy.

Our review of the hospital case notes showed that in as many as one quarter of patients in the trial there was disparity within the trial records. Most of the disagreements were minor and did not influence the outcome of the analysis. However, there were 4 patients in whom the protocol of treatment policy was not obeyed, in addition to the three patients we had previously reported; in 9 patients there was disagreement over the site of first recurrence, and in 8 patients disagreement over their current status, (i.e. whether they were free of recurrence or not). In other patients the date of mastectomy, the date of first recurrence or date of death was different by a matter of a few days. In 16 patients however, the date of first recurrence was significantly different from the trial record, and this was because

	Simple Mx			Radical Mx	
	Node Negative	Node not Identified	Given XRT	Node Negative	Node not Identified
Total in Group	37	26	2	60	2
Recurrence:					
Scar	5	3	1	10	0
Nodes:Axillary	6	5	1	1	0
Supraclavicular	2	3	1	4	0
Internal Mammary	0	1	0	1	0
Total No. of Patients with Local Recurrence	9 (24%)	8 (31%)	2	14 (23%)	

TABLE VI

Failure of local control (analysed April 1978) of the disease, in patients with no evidence of spread to the axilla at the time of mastectomy. Note that patients treated incorrectly according to trial protocol are recorded separately according to their node status.

	Simple Mx		Radical Mx	
	+ XRT	no XRT	+ XRT	no XRT
Total in Group	37	1	31	4
Recurrence:				
Scar	11	0	1	0
nodes				
axillary	0	1	2	0
supraclavicular	1	0	2	1
internal mammary	0	0	0	0
Total No. of patients with local recurrence	11 (30%)	1	4 (13%)	1

TABLE VII

Failure of local control (analysed April 1978) in patients with proven tumour spread to the axillary nodes at the time of mastectomy. Patients treated incorrectly according to protocol are recorded separately.

of different opinions about its definition.

Conclusions

From the analysis of the trial data, we feel that a conservative policy based on the histological examination of pectoral nodes does not impair survival or the incidence of disseminated disease. In patients with spread of tumour to the axilla, radical radiotherapy protects against chest wall recurrence and axillary radiotherapy alone should not be practised. There is no doubt that physical morbidity increases with the extent of treatment.

We have learned that the definitions of the trial when it commenced 10 years ago were not as specific as they might have been. We would now define our entry criteria more carefully and insist on stricter adherence to operative techniques.

A consistent definition of "recurrence" is essential, and reports of trials should indicate whether the analysis is of first recurrence or all recurrences. As trials of local therapy are likely to show differences only in local control of disease, attention should be given to local failure rates. This is important even though survival may not be compromised.

Finally, there is no doubt that regular external review of trial data is essential.

References

1. Forrest, A.P.M., Gleave, E.N., Roberts, M.M., Henk, J.M., Gravelle, I.H.
Proc.Roy.Soc.Med., 63, 107 (1970

2. Roberts, M.M., Forrest, A.P.M., Blumgart, L.H., Campbell, H. et al.
Lancet, i, 1073, (1973)

3. McNair, T.J., Dudley, H.A.F.,
Lancet, i, 713 (1960)

4. Forrest, A.P.M., Roberts, M.M., Cant, E.L.M., Shivas, A.A.,
Brit.J.Surg., 63, 569, (1976)

5. Peto, R., Pike, M.C.
Biometrics, 29, 579 (1973)

Comments to Dr. Roberts' paper

Dr. Stewart

I would like to stress the importance as well as the difficulty in timing the failures (e.g. recurrent disease) in trials of primary breast cancer. Rules have to be defined and adhered to, signs of disease rather than symptoms have to be timed and the dating often requires to be corrected in respect of subsequent behaviour of disease. Subsequent external review is therefore advised prior to trial analysis. This was done for the Cardiff trial and resulted, through discussion of alterations, in the more accurate recording of the results of treatment options.

Experiences from a Multicentric Trial of Adjuvant Chemotherapy

Roar Nissen-Meyer
Oslo, Norway

Hypothesis and aim of the study

The prognosis after an intended curative operation for a primary cancer includes two factors:
A: The percent of the patients definitely cured, and
B: The speed of disease progression in the cases not cured, - measures by the time until clinical recurrences and by the survival time.

Both these two factors of the prognosis may be improved by adjuvant chemotherapy, but we should differentiate clearly between the two types of effect. They also need different statistical methods for testing the significance of the results.

From animal experiments we know that it is essential for obtaining an effect of type A, killing of the last cancer cell, that only single cancer cells or clusters of few such cells are present in the animal. In such a situation it may be sufficient with one single but intense chemotherapy course, given as soon as possible after dissemination of the cancer cells.

On the other hand, for obtaining an effect of type B, a reduced growth rate of remaining tumour tissue, there must be a continuous influence of the growth reducing factor over a long time.

The situation with only single cancer cells or clusters with a few such cells together is easily produced in animal models. Under clinical conditions in humans, however, such a situation is probably only met immediately after surgical removal of the primary tumour, and provided the first dissemination of cancer cells from this tumour took place during or immediately before this surgery.

How often this happens we do not know, but we have reason to assume that most often the first dissemination takes place long before diagnosis of the primary tumour, often years before. In such cases the subclinical metastases will at the time of surgery contain too many cells to be totally eradicated by chemotherapy and the best we can hope for is to delay their growth.

The hope for extra cures by adjuvant chemotherapy lies in the possibility that in some cases the first dissemination took place during or

immediately before surgery. To explore this possibility was the aim of the present study.

An effect of type A, an increased cure rate, will logically be seen in the survival curves and the curves for diseasefree survival of a controlled clinical trial as increasing and then persisting difference between the treated and the control group.

An effect of type B, however, will produce delay curves, with the largest difference in the beginning, but then the curves will approach each other again.

Combinations of the two types of effect may of course also occur. In our study we gave only one single six-day chemotherapy course, and could expect that no major delay effect would obscure the hoped-for effect of type A.

The type of effect we were looking for made us prepared for a very long follow-up period, and especially interested in the difference between the treated and the control group during the later years of this follow-up period. We were not particularly interested in sophisticated statistical methods enabling us to detect and publish a significant difference between the two groups early, before we could see the type of the difference.

Moreover, since we could expect to obtain only small differences in the longterm results, we had to build up a large series, in order to have enough cases left for statistical evaluation of the late recurrences. Such a series was only possible to obtain by a multicentric trial, in areas with a stable population.

Performance of the study

A number of cancer treating hospitals in Scandinavia agreed to join in a multicentre trial according to these principles. On October 10th 1964 our committee designed the principal protocol. The design of the study had to be simple and easy to follow, so as not to interfere too much with the routine work in these busy hospitals. Each hospital had to standardize and define its methods for surgery, postoperative radiotherapy and postoperative endocrine therapy, and to stick to these methods for all cases found eligible for this study. Variation in treatment methods were allowed between hospitals, but not within hospitals.

It soon became evident that it was difficult to get all parts involved in a hospital to agree about the stadardization of treatment methods

for a future of about 5 years. Some hospitals had to wait for the appointment of a new head of one of the departments to discuss the matter with him. In some hospitals new equipment had been ordered for the radiological treatment, and they had to wait for this to be implemented.

As a result of these practical obstacles, some of the originally involved hospitals never became active in the trial, but new hospitals joined our group when it was evident that our organization did work. A total of eleven hospitals in Finland, Norway and Sweden eventually started randomizing patients according to our protocol, the first one on January 15th 1965, the last one in December 1968. In 1973 the committee felt the trend of the results in favour of the treatment group to be so convincing that it recommended the randomization to be stopped. Some of the participating hospitals, however, wanted to continue, in order to reach a reasonable size of their own part of the series. The last patient was randomized September 25th 1975. The follow-up study is still continuing.

Protocol violations

A total of 1188 patients were randomized, 586 to the treatment group, 602 to the control group. Of these, however, 4.38% (27 from the treatment group and 25 from the control group) were excluded due to ineligibility. In 25 of these cases the first diagnosis, usually made on the basis of a frozen section, was wrong. The final histologic diagnosis was benign disease in 22, malignant disease other than mammary cardinoma in 3 cases. In 27 cases the busy surgeon had overlooked a history of a previous malignant disease or positive x-ray findings, before surgery and randomization took place. We can probably never totally avoid such mistakes if the treatment group are to be treated immediately after surgery. These cases were withdrawn from the study, and had no influence on the results.

By mistakes in the wards 8 patients of the control group were given chemotherapy, and 11 of the treatment group did not receive the chemotherapy they were allotted to. In addition the dose of Cyclophosphamide was too low in 27 cases, for one reason or other. These patients are <u>not</u> excluded from the analysis of results. - In summary, 4% of the patients did not receive exactly the treatment they should have. By keeping them in the group they were randomized too, the validity of the study is not much influenced, - there will only be a small dilution of the groups.

Randomization procedure and distribution of variables

Each hospital had its separate protocol for blind randomization, to which the local investigator had no access. It was kept by a secretary in another department, and entry of new cases was made over the telephone.

Since we were building up a really large series, stratified randomization according to some of the known prognostic factors was found unnecessary. The large number of cases would secure a random distribution of all the prognostic factors, both the known and the unknown, and allow detailed analysis after retrospective stratification. A conventional stratified randomization is also not possible in this type of study, where the final histology is not available before randomization. One has to rely on the random distribution in the large series.

A check on some of the variables confirms this random distribution between the treatment group and the control group, e.g. premenopausal patients 34.5 and 35.9%, nodes positive 41 and 42.3% respectively.

Processing of data

All pertinent data are transferred from the usual medical records to large, manually operated punch cards. These are completed at my yearly visit to the hospitals, and make the subdividing of the case material (retrospective stratification) and the construction of life tables very easy, without technical equipment. A cheap electronic pocket calculator is enough for the statistical tests.

Expenses

Treatment in hospital is free for all patients in the Scandinavian countries, and the doctors did their part of the study work as part of the routine job they were paid for by the hospitals. The Cyclophosphamide used for the treatment group was bought over the regular hospital budget, just as all other drugs used in the hospitals, - the cost was less than 40 US$ per patient. Our only direct expenses were for printing of the punch cards and reporting forms, and for travelling, and these expenses were generously paid by AB Pharmacia in Uppsala, the manufacturer of the Cyclophosphamide used.

Follow-up

When more than 5 years had elapsed since surgery, the patients were due for routine control only once a year. Once a year, too, I visit the participating hospitals for updating of the punch cards. Theoreti-

cally, then, every living patient should have been seen less than one year previous to my visit. For several reasons, however, the control could have been delayed for a short time for a few patients. All these appeared shortly afterwards when they were called, and the delay has not caused any statistical problems.

A problem, however, is caused by 30 patients who have disappeared from our control for more than 2 years, and who must be considered as lost to follow-up. The majority of these lost patients (26 cases) came from one hospital. This hospital had difficult working conditions, serving an immense area with a widely scattered population. Some years ago the organization of the follow-up from this hospital virtually collapsed. We discussed if all patients from this hospital should be excluded from the study, since these 26 lost patients represented 38.8% of its total case material. We decided to keep them in the study, and only register the lost patients as "withdrawn alive", for the following reason: The hospital had an even distribution of patients, 31 vs. 31, and the numbers of recurrences and deaths recorded were also evenly distributed. An exclusion of the total material from this hospital would therefore mean an improvement of our results. However, we have learned that in our next study we will consider carefully if a hospital wanting to join us really has a fair chance to complete the study satisfactory, including more than 10 years follow-up.

The other 4 patients missing either emigrated, or wanted for religious reasons not to have any more to do with doctors.

Evaluation of end-points

The results have been presented both as percentages free from recurrent disease and percentages crude survival. As recurrent disease has been reckoned diagnosis of distant metastases, local recurrences and cancer in the opposite breast (the latter for practical reasons, since it can not always be distinguished from a metastasis).

We met one quite important problem, the magnitude of which we had not realized on beforehand: patients dying under conditions suggesting metastases, but without definite proof of recurrent disease, e.g. younger, apparently healthy women dying from 'cerebral hemorrhage', which might suggest cerebral metastases, or with icterus, suggesting liver metastases. It was depressing to see how often old, debilitated people could die in their own homes or in institutions under the diagnosis of an intractable abdominal cancer, without efforts made to secure the nature of this cancer. It might have been a new abdominal

cancer of some kind, but in some cases it might have been metastases from the breast cancer, which could perhaps have been palliated by endocrine treatment.

We discussed how to handle this problem statistically, and decided to look at the randomization group when we were in doubt if recurrent disease should be registered. If the patient belonged to the control group, we would be very reluctant to register a recurrence, if she belonged to the treatment group we would rather accept the recurrence diagnosis. In this way we would at least not make the results of our adjuvant chemotherapy appear better than they should.

There is no such dispute about life or death, and we always got the exact date of death. With increasing follow-up time the crude survival rate may therefore be considered a better measure of results than recurrence rate. We notice that we have now a greater difference in death rate than in recurrence rate - probably because of our cautiousness about doubtful recurrences discussed above.

However, the patients being 65 - 70 years at surgery will after 15 years of follow-up be 80 - 85 years, and the non-cancer deaths will gradually dominate and dilute the picture. One way of solving this problem may be to register the patient in the column 'withdrawn alive' in the life table as soon as she reaches the age of e.g. 80 years, or perhaps 75. This has not been done yet, but we will consider it in a later updating of the results.

Statistical evaluation

The actuarial or life table method has consequently been used, with estimation of significance according to the method described by Cutler & Ederer 1958. We still feel this to be a useful method for our purpose, while newer methods like the log rank test would probably be better if we had been more interested in early detection of a delay effect of the treatment.

Two principally different subseries

The original aim was to study the effect of a short chemotherapy course given immediately after surgery. In ten out of the eleven participating hospitals randomization took place in the surgical department, and the chemotherapy course started immediately after completion of the surgical procedure. The patients from these ten hospitals constitute our main series of 1026 cases.

However, we also wanted to study the significance of the time interval between surgery and the adjuvant chemotherapy course, since this played such an important role in the animal experiments we had as a basis for our hypothesis. A delay of a few days could seriously reduce the effect on ultimate cure rate in these experiments. We therefore took advantage of the possibility provided us by the following circumstances: One of the hospitals willing to participate in our study had working conditions quite different from the others. It was a Radiotherapy institute, and the mastectomy was performed in other hospitals. When the patients came to the institute for randomization and treatment, between 2 and 4 weeks had elapsed since surgery. By keeping this routine unchanged, we could establish a subseries of 110 cases where the chemotherapy course was given at an average 3 weeks after surgery, - but the chemotherapy course in itself and the rest of treatment was the same in these two subseries.

Results

The treatment group received one single course of Cyclophosphamide, with 5 mg per kg daily for six days only. This treatment carried no risk, and did not interfere with the routine surgical and radiological treatment. The side effects were mild and of short duration.

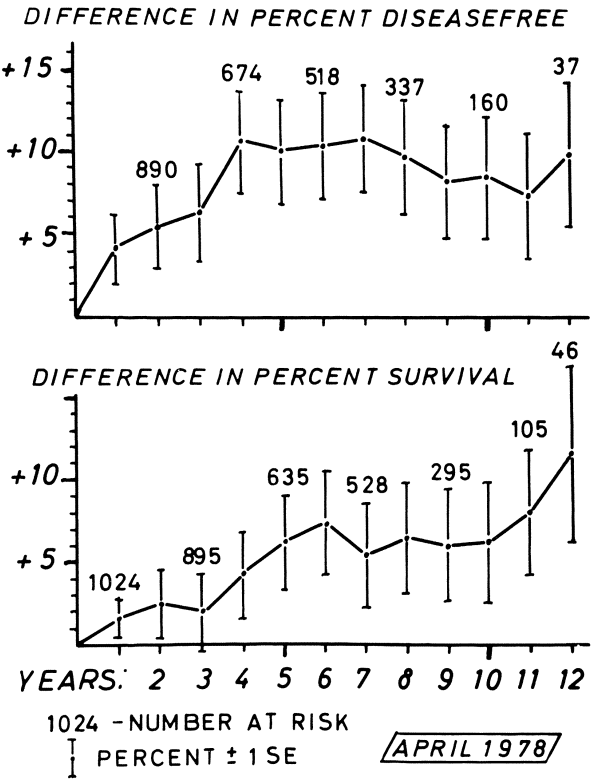
The chemotherapy group has a total of 559 cases, with 232 recurrences and 210 deaths, the control group 577 cases with 284 recurrences and 265 deaths. The differences of 52 recurrences and 55 deaths are both significant with $P < 0.01$.

In the main subseries (Fig. 1) the difference in percent diseasefree increased until 4 years after mastectomy, and was thereafter kept about 10 percent. The difference in percent crude survival had a slower increase, and has reached 11.5% 12 years after mastectomy. This seems well in accordance with our hypothesis and the animal experiments. We have probably increased the ultimate cure rate with about 10 percent.

We cannot, however claim any effect in the smaller subseries with a delay of 2-4 weeks until chemotherapy. We believe that 3 weeks after surgery we have passed the optimal time for obtaining cure by adjuvant chemotherapy (but of course it will not be too late to obtain a delay with a long-term chemotherapy).

Regardless how we have tried to stratify our main series, if only the groups are of a reasonable size (near 100 or more), the results show the same trend as demonstrated in Fig. 1, but due to the smaller number of cases it is more difficult to establish the statistical signi-

Fig. 1 Results in the main subseries of the Scandinavian Adjuvant Chemotherapy Study Group



ficance. Especially at the beginning of the curves the number will be so small that there is a greater variation due to chance. The results have been presented in greater detail in previous publications (Nissen-Meyer et al. 1978) and shall not be repeated here.

Conclusions and way ahead

We have reached the following conclusions:
One single, short adjuvant chemotherapy course may increase the ultimate cure rate, as it has been obtained in animal models.

It seems essential that this chemotherapy course is given immediately after surgery, as it has been shown in animal models.

We find benefit also in the prognostically most favourable groups of patients.

We find benefit in postmenopausal and in primarily castrated groups of patients as well as in premenopausal patients with active ovaries.

Extension of adjuvant chemotherapy to intense treatment for one year or more will in addition introduce a factor of delay, but will also cause side effects and considerable distress for the patient during the time of treatment. We do not know enough about the long-term risk and side effects of such treatment.

The Scandinavian Adjuvant Chemotherapy Study Group started in March 1977 a new series to compare directly the two principles of adjuvant chemotherapy: We give all patients one short chemotherapy course immediately after mastectomy, and one half of the node positive patients continue with CMF courses monthly for one year.

References

Cutler, S.J. and Ederer, F.: Maximum utilization of the life table method in analyzing survival.
J.Chronic Dis. 8 (1958), 699 - 712.

Nissen-Meyer, R., Kjellgren, K., Malmio, K., Månsson, B. and Norin, T.: Surgical adjuvant chemotherapy. Results with one short course with Cyclophosphamide after mastectomy for breast cancer.
Cancer 41 (1978) 2088 - 2098.

Comment to Dr. Nissen-Meyer's paper

1

Dr. Prentice

Dr. Nissen-Meyer as well as some other speakers presented results that suggested treatment effects, in this case corresponding to adjuvant chemotherapy, that appear most pronounced early and somewhat less so later in the patients' course. In technical terms the ratio of mortality rates in the two treatment programs is apparently larger soon after the beginning of the trial than subsequently. I should like to point out that alternative tests to the so-called log-rank test, such as the generalized Wilcoxon test may be better able to detect such treatment effects. To put it another way, if one suspects in advance that the treatment differences are likely to be most pronounced early in the study, then a smaller number of patients would usually be required to test this hypothesis if a generalized Wilcoxon, rather than a log-rank test is used.

Thirty Years Experience with Breast Cancer Clinical Trials at the Christie Hospital in Manchester
- Clinical Aspects -

G.C. Ribeiro

Introduction: In this paper I will deal with the clinical aspects of the Manchester trials. In his paper Dr. Palmer will describe the statistical lessons learned in particular discussing the number of patients required in clinical trials, the recording of information and the analysis and follow-up of these patients.

The first trial carried out between 1949 and 1955 at the Christie Hospital was designed to compare the results of X-ray therapy given shortly after a radical Halsted mastectomy (treated group) with those of X-ray therapy delayed until some local recurrence demanded it. (watched group). All patients were operable and there had to be no doubt in the surgeons mind whether there was any residual disease. All patients were under 70 years of age and there was histological evidence in all cases.

Two radiotherapy techniques were used in the treated group and these are referred to as the quadrate and peripheral techniques. A total of 1.141 patients were entered into the trial.

At 15 years there was no significant difference between the survival of the watched and treated groups in the quadrate series. If the patients are split (histologically) into those with positive nodes in the axilla and those with negative nodes there is again no significant difference in survival between the various groups. For those with positive nodes 22% of the treated group survived 15 years compared with 20% of the watched patients. In the patients with negative nodes 45% of the treated group survived 15 years compared with 52% of the watched group.

Very similar findings were obtained for the peripheral series, where again no significant difference in survival was found in the treated and watched groups.

Although it is now obvious that prophylactic irradiation did not influence survival, it significantly reduced the incidence of local recurrence. The incidence of local recurrence was 32% in the watched groups compared to 19% in the treated patients. Furthermore the watched patients who were subsequently irradiated for local recurrence

had local control achieved to the same degree, so that at death, local recurrence was present in only 16% of the original watched patients.

35% of the watched patients developed distant metastases without evidence of local disease and were spared the discomfort of radiation; it could be that 35% of the treated group of patients were also given post-operative therapy needlessly, as they may never have got recurcence of local disease.

It was unfortunately mentioned in an earlier report of Paterson and Russell that there was an excess of liver metastases in the patients given radio-therapy and this fact has frequently been quoted against radiotherapy. In fact this was noted in the quadrate series only and in the first year after radiotherapy. The number of cases were very small, less than 3% of the series, and at 7 years there was no difference in the incidence of liver metastases between the treated and the watched groups.

A second trial was also carried out between 1948 and 1955. All the patients had operable breast carcinoma and had had a radical mastectomy. They were then randomised into two groups; one group had an X-ray artificial menopause(irradiated) and the other did not (controls). All pre-menopausal patients were included up to 2 years past the natural menopause. A single X-ray treatment was given delivering 450 rads to the mid-pelvis. The object of the trial was to compare survival rates and recurrence/metastatic rates between radiated and control groups.

At 10 years 54.9% of the radiated patients were alive and well compared to 47.5% of the controls (crude survival rates, p=0.07).

It is also of interest to note the cumulative incidence of recurrence. At 10 years, 47% of the irradiated group had developed recurrence/metastases compared to 48% at 4 years in the control group and at 15 years 51% of the radiated group had recurrence compared to 52% at 7 years in the controls. When analysed by age, the patients aged less than 40 years benefitted least and their survival was worse than that of the older groups both in the radiated and control groups. The return of menstrual bleeding was also highest in these young patients but the survival of those patients who had a return of periods was no worse than those whose periods were stopped with the one X-ray treatment.

Since 1955, patients under 40 years of age usually have an oophorectomy in preference to X-ray therapy . If X-ray therapy is used however then the dose is now 1500 rads in 4 treatments on megavoltage. It is also the practice now not to routinely do a prophylactic menopause on operable patients under 35 years of age.

In the last decade the surgical approach to early, operable breast carcinoma has changed considerably in the United Kingdom, with simple mastectomy finding far more favour among surgeouns than radical mastectomy. In addition, during this same period papers appeared suggesting that radiotherapy was harmful and strongly suggested that axillary nodes that were not involved with malignancy should not be irradiated as this would interfere with host immunological mechanisms.

In view of these suggestions, a further clinical trial was carried in the Manchester region between 1970 and 1975.

Patients with Clinical Stage I and II breast carcinoma were eligible (UICC 1968). A total of 1020 patients were entered. Patients with Stage I disease were randomly allocated to having either a simple mastectomy alone or simple mastectomy with radiotherapy postoperatively (713 patients).

307 patients with Stage II disease were allocated to either having a simple mastectomy with radiotherapy or a radical mastectomy alone. In addition, all premenopausal patients were offered an artificial menopause by X-ray therapy or oophorectomy. Preliminary results have been published and these show that there is no significant difference in survival at 5 years between the various groups, but that once again radiotherapy is shown to significantly reduce local recurrence, following simple mastectomy. It has also reconfirmed the view held by us that radiotherapy given to axillary node negative patients is not harmful and does not adversely affect their survival.

Since November 1976, we have been entering patients into the latest of the trials to be held at the Christie Hospital. Patients with operable tumours (T1, T2, T3a, NO and N1, MO) are randomised following surgery, and where necessary given radiotherapy. Premenopausal women are allocated to either an artificial menopause or Tamoxifen and postmenopausal women to either Tamoxifen or no further treatment (controls). The drug Tamoxifen is given in a dosage of 10 mg twice daily for one year.

In conclusion, we feel that valuable clinical knowledge has been gained over the past thirty years and that it has been done in a valid scientific manner. The trials continue in a logical manner and it is merely a matter of time before significant advances will be made in the improvement of survival in breast carcinoma.

References

-1 Clinical trials in Malignant Disease. Paterson & Russell. Journal of Faculty of Radiologists 10 (1959) 175-180.

-2 Manchester Regional Breast Study
J.P. Lythgoe et al. Lancet. (1978) 744-747

Comments to Dr. Ribeiro's paper

1

Dr. Wallgren

The early trials conducted in Manchester were not true randomized trials. The patients were allocated to various treatments by means ot the date of birth. Since the treatment in the particular case was known in advance this is likely to have resulted in a biased selection of cases in the two treatment groups. A visible example of this is the different number of cases in the so-called 'quadrate' series of radiation therapy. Fewer patients were allocated to the radiotherapy group (327) than to the watch policy group (393; $p \approx 0.05$). This may well have biased the results.

2

Dr. Roberts

The local recurrence rate in the Manchester series* is lower than in the Cardiff series. The reason for this is unclear, but it may be because the Manchester trial has only just reached 5 years follow-up.

* Second trial (1970-75) [Editors' note]

Thirty Years Experience of Breast Cancer Trials in Manchester England - Statistical Aspects.

Michael K. Palmer Ph.D.

Other speakers have described some of the statistical aspects of clinical trials in breast cancer in more detail than I am able to do in the time available to me, but I do hope I will be able to pull together some of the improtant points which have already been mentioned.

The first principal is that, for a trial to be useful, large numbers of patients must be entered. It is realistic to assume that there will not be dramatic improvements in the treatment of breast cancer in the foreseeable future. There is now overwhelming evidence that, in most cases breast cancer is a systemic disease virtually from the outset so many patients will have undetected micro-metastases at diagnosis and will not be cured by purely local forms of treatment. Therefore local treatments will be either identical in their effects, or the differences between them will be very small. The numbers of patients in trials must therefore be very large. One cannot be dogmatic about the minimum number of patients required in a trial but power considerations indicate that one should probably aim for at least 250 on each side. Of course the problem is that very long trials tend to be overtaken by events. The reasons for comparing two particular treatments may have been very good in 1968, but in 1978 those reasons will probably appear inadequate. So the total patient intake should ideally be achieved in as short a time as possible, say 3 years. This means that most trials will have to be multicentre if the two conflicting aims of large numbers and short durations are to be achieved. I would suggest that a trial with a few large centres would be much better than one with a lot of centres contributing only a small number of patients.

Now, turning to the question of randomisation, a trial with large numbers of patients usually makes stratification unnecessary, although in a multicentre trial, for the peace of mind of the participants, one would want to stratify by institution as was done in the Scandinavian trial described by Dr. Nissen-Meyer. So a straight-forward balanced randomisation will produce two virtually identical groups. The effects of chance differences between groups in the distributions of prognostic factors can be corrected using modern methods of statistical analysis. I will say more about this later. We have found that randomisation can be achieved best by telephone call to a central

office which holds a pre-prepared, balanced randomisation, and this office can also monitor trials continuously. So the second vital principle is that there should be a central office responsible for randomisation, documentation, monitoring and eventually, of course analysis.

Good documentation is vital. Forms should be designed which are comprehensive yet not too laborious to complete. The design of suitable forms takes a lot of time and effort but their importance cannot be overstressed. Typically there may be two or three versions before the trial even starts. And probably the first few patients in the trial will throw up some inadequancies which require further changes. All prognostic factors must be recorded and I would suggest the following as a minimum.

1. TNM classification.
2. Histological status of axillary nodes.
3. Menopausal status.
4. Site within the breast.
5. Histological features of the primary tumour.

At the risk of over-simplifying complex relationships, let me say a few words about each of these.

A primary tumour less than 2 cms is designated T1 in the UICC 1968 classification, while if it is more than 2 cms but less than 5 cms it is designated T2. Figure 1 shows the substantial difference in survival between these two T-categories that we found in one of the Manchester Trials (reported by Lythgoe et al. 1978). Everyone participating in a trial would have to be clear about the measurement of tumour diameter, whether it was a clinical or a pathological measurement, and whether it was the maximum diameter of the tumour. Dr. Baessler described some of the problems in measurements of this type in his lecture. NO in the UICC classification means that nodes are not palpable in the axilla while N1 means that they are. Providing the primary tumour is T1 or T2, this defines two clinical stages and figure 2 from the same Manchester trial shows a clear difference in survival.

Histology of the axillary nodes: it is not enough just to say that the axilla was positive. Fisher, Slack and their co-workers (1969) found a clear difference in prognosis depending on whether 1-3 or 4 or more nodes were involved.

Table 1 shows five year survival percentages of 76% for patients with axilla negative, 62% for patients with one to three positive nodes in

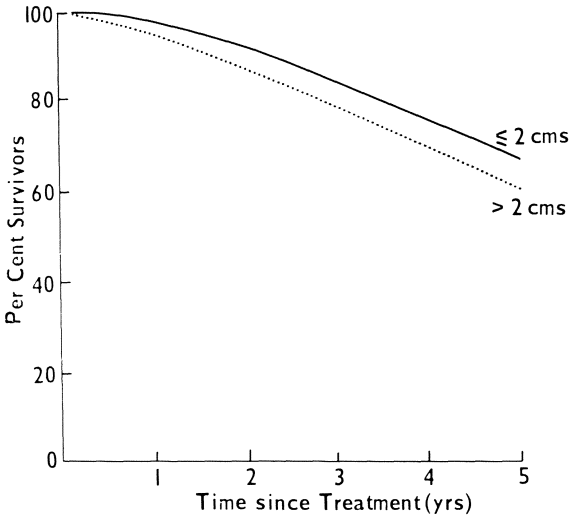


Figure 1 Manchester Regional Breast Study: Survival related to Maximum Tumour Diameter

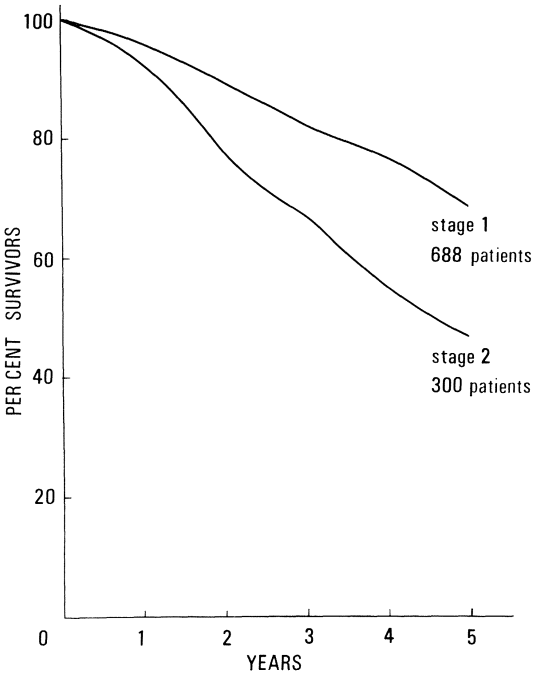


Figure 2 Manchester Regional Breast Study: Survival related to Clinical Stage

the axilla and 31% for patients with four or more positive nodes.

Table 1.

Effects of Nodal Status on 5 year Survival Percentages

Nodal Status	Patients	% 5 yr. survivors
Negative	539	76
1-3 positive nodes	281	62
4+ positive nodes	285	31

Fisher, Slack and Bross. Cancer 24 1071 - 1080.

This prognostic information can be provided only if the maximum possible number of nodes are removed at operation, and the number of nodes histologically positive and the total number examined should both be recorded. Other pathological features are relevant, such as histological grade, degree of cellularity, and lmphocytic infiltration. Ideally, all these investigations should be carried out by the same pathologist, but this would be a major undertaking. The results of oestrogen and progesterone studies should also be recorded, at least in a sample of patients, to identify hormone sensitivity. Site of tumour within the breast appears to influence the probability of involvement of various regional lymph nodes, although in itself it does not seem to be a prognostic factor. Nevertheless, site within the breast should probably be recorded.

In most series including the Manchester Trials, menopausal status has been found to influence prognosis. Three groups can be identified: premenopausal women, women who had a recent natural menopause, say within the previous three years, and women who had a menopause more than three years ago. It is interesting to note from figure 3 that it is the women who had a menopause most recently who have the worst prognosis.

Let me now turn to the question of analysis, although I know other speakers will have more to say on this subject. I want to consider for a moment the cost of statistical analysis compared with other costs associated with a breast cancer trial. In Britain at the present time the medical cost of treating a woman with breast cancer in a big

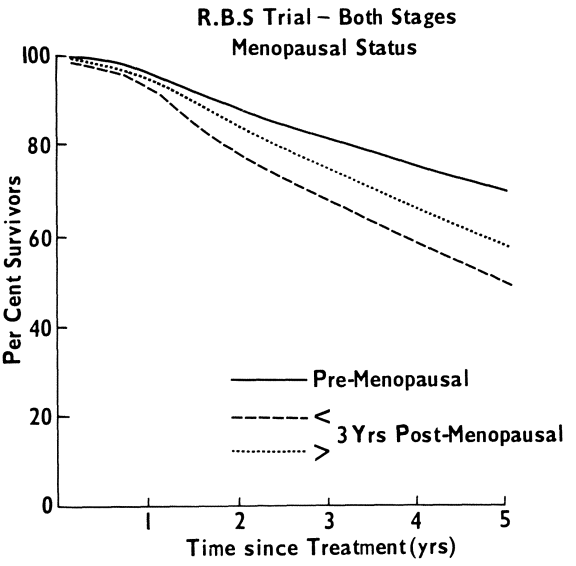


Figure 3

Manchester Regional Breast Study: Survival related to Menopausal Status

centre is about £ 400 for the surgery, plus a further £ 800 for a course of post-operative radiotherapy if this is required. If all other costs, such as convalescence, time off work, community help, social security payments and so on are ignored, the total medical investment in a breast cancer trial in which 500 patients were randomised and received both surgery ans X-ray treatment would be in the region of £ 600.000. If patients also received adjuvant chemotherapy, the total cost would be even more. Of course this amount of money would be spent whether or not the patients were in a clinical trial, but the point I am making is that the statistical analysis of the results of a trial must be detailed, and even the most painstaking and comprehensive analysis would not cost more than 2 or 3% of the medical investment in the trial.

The features that ought to be analysed include survival, disease-free survival, freedom from local recurrence and freedom from distant metastases. The four possible types of failure, not mutually exclusive, for each patient are therefore: death, recurrence of disease anywhere, recurrence locally, (that is in the breast or in the axillary or supraclavicular lymph nodes), and the appearance of disease in distant sites. The times of occurrence, or non-occurrence, of each of these events must be documented separately and carefully, and life table graphs constructed in the usual way.

Patients dying from intercurrent illnesses, apparently free from all evidence of recurrence of breast cancer, can cause problems in the analysis. In some trials which have been published in the medical literature these patients have been completely excluded from the analysis, in others no distinction has been made between deaths from breast cancer and deaths from other causes. In fact the proportion of intercurrent deaths is usually so small that it doesn't matter, but we think there are advantages in a uniform approach to this. Our opinion is that the results for these patients should be retained in the analysis and their times should be censored at the point of death. However the cause of death must be completely unrelated to the breast cancer - indeed it may even be a primary in the other breast. If there is any doubt whatsoever as to whether the cause of death was in any way related to the breast cancer for which the patient entered the trial, then it should be assumed that it was. The criteria for intercurrent death are therefore very strict, and this policy means that patients must be kept on close clinical follow-up, so that events such as recurrence of disease locally, and the development of distant metastases can be documented accurately. We feel it is not satisfactory

to rely just on death certificates for the cause of death since these are known to be inaccurate in many cases.

As far as presentation of results is concerned, we think it is important to show life table curves since these give much more information and a better impression of the results than just say, the 3 year or 5 year percentages, although the latter should also be stated in a publication so that more exact comparison can be made with the results of other trials. The results of treatments must of course be compared within subgroups of patients defined by the different levels of each prognostic factor. The question implicit in clinical trials used to be 'which of the two treatments is better overall?' With modern methods of statistical analysis we can now answer the question: 'which type of treatment is better for certain types of patients?' and this is much more useful and relevant. So treatment differences must be explored within subgroups of patients and this is possible in a large trial. Treatment effects must also be refined to take into account differences between treatment groups in the distribution of prognostic factors, although this is expected to have little effect in a large trial. I might mention at this point how very useful we find Richard Peto's survival analysis program: this calculates and plots lifetable curves for different values of covariates, either for all patients or for subgroups of patients. It also carries out logrank tests including adjustment for other prognostic factors. Regression methods are also of tremendous value, but I won't mention these since more time will be devoted to them in the next few days.

Survival and recurrence are not the only measures of the success of treatment however. If treatments give very similar prognoses then other measures of response assume greater importance. So we should not forget to consider functional results, that is such things as arm swelling and shoulder stiffness; cosmetic results that is how much of the breast and surrounding structures have been removed; and treatment toxicity, from radiotherapy, hormones and particularly from adjuvant chemotherapy. Psychological aspects are also important, and Maguire et al. (1978) assessed anxiety, depression, guilt and stress in women undergoing breast surgery. He compared a group of breast cancer patients with a group who had biopsy for benign breast disease. Psychiatric morbidity, anxiety, depression and sexual problems, were significantly higher in breast cancer patients. It is important to emphasise that these were not trivial symptoms. They were moderate or severe symptoms, recognised by a psychiatrist as requiring psychiatric treatment.

With detailed information like this, it may eventually be possible to estimate in some way the cost/benefit ratios for particular treatments; the cost being defined as the disability, toxicity and psychological effects associated with a given treatment, and benefit as disease-free interval or survival time.

Now, with regard to publication of the results of breast cancer trials, it seems inevitable that statistical analyses will be carried out on several occasions, and results will be published on at least three. The first time, the preliminary results would probably be presented for the first 200 or so patients in the trial, then a report would appear soon after the termination of patient entry and then again after a full five years follow-up on all patients. In addition, it is clearly desirable eventually to publish the long-term results when all patients have matured to at least ten or fifteen years, in order to assess the very long-term effects of treatment. Perhaps, in view of this, one should adopt more stringent significance levels, or even use fully sequential methods, such as the sequential logrank test. Some support for more stringent significance levels comes from the realisation that with perhaps 100 breast cancer trials being carried out world-wide at any single time, and in probably 98 of these, treatments are virtually equally effective, then one would expect about 5 'significant' results by chance alone. And, if results of trials are to be analysed in the sort of detail I suggest then the number of 'positive' findings must be further increased.

To conclude, then, I want to say that we now have very sensitive and powerful statistical techniques for the analysis of clinical trials in breast cancer. These trials should be carefully designed, and meticulously carried out. Attention should be given to recording all prognostic variables, and they should be analysed and reported in a uniform way, so far as this is possible. Above all they should be large. This requires on the part of surgeons, radiotherapists and oncologists that they collaborate, nationally and internationally, to produce good, carefully worked out and simple protocols, and new concepts for the treatment of breast cancer.

References

Fisher, B., Slack, N.H., and Bross, I.D.J.: Cancer, 1969, 24,1071-1080

Lythgoe, J.P., Leck, I., and Swindell, R.: Lancet, 1978, 1, 744-747.

Maguire, G.P., Lee, E.G., Bevington, D.J., Kuchemann, C.S., Crabtree, R.J., and Cornell, C.E.,:B.M.J. 1978, 1 963-965

Comments to Dr. Palmer's paper

1

Dr. Roberts

We measured psychological morbidity in a number of patients in the Cardiff trial, and found that it was similar in patients who had the radical policy and those treated conservatively. On the other hand, physical morbidity measured by degree of arm elevation and arm oedema was significantly increased in patients who had radical treatment compared with conservatively treated patients.

Answer to Dr. Roberts

2

Dr. Palmer

I am grateful to Dr. Roberts for her remarks concerning the Cardiff breast cancer trial. Her findings emphasise the importance of taking into account the quality of life of patients with breast cancer when assessing the relative merits of alternative treatments, particularly when there are no differences in the survival distributions.

3

Dr. Lagakos

Although I heartily agree with most of Dr. Palmer's comments, I must express my disagreement with his opinion that stratification (at the time of randomization) is unnecessary in multi-institutional studies. If the randomization must be carried out using the closed-envelope technique, thenperhaps a case can be made against stratification on the basis of simplicity. However, when the randomizations are carried out at a central facility, they should always be stratified to ensure treatment balance across institutions and with respect to the important prognostic factors rather than leaving this balance to 'chance'.

It is important that treatment-groups be balanced for several reasons: (1) The power of statistical tests comparing treatments is highest when there are no imbalances; (2) Unadjusted treatment comparisons for subgroups of patients will not be biased as a result of imbalances in one or more of the prognostic factors [with a simple(i.e., unstratified) randomization, these comparisons will only be unbiased 'on the

average'. Loosely speaking, this means that if the experiment were repeated many times, for every experiment where there was a bias of a given direction and magnitude there would correspond another experiment with a bias of equal magnitude but in the opposite direction. Over all experiments, these biases would 'average' to zero. However, in any particular experiment they would in general be nonzero]; (3) Biases due to institutional differences that cannot be explained by known prognostic factors [e.g.,differences in patient populations, differences in the evaluation of subjective criteria such as Karnovsky Status, differences in supportive therapy etc.] will be avoided by balancing treatments within institutions.

The cost of stratification is small: at the time of randomization, one need only ascertain the patient's prognostic status and then use a randomization book that has sections corresponding to the various levels of the prognostic factors. Our institution has used this system for years on a great number of clinical trials and I can assure you that stratification can be accomplished with virtually no additional effort or complications.

Some people argue that the advantages of stratification are 'on the average' small. To these people, I can only respond that it makes no sense to leave to chance that which can be assured with virtually no additional effort.

4

Dr. Tsiatis

Dr. Palmer made the comment that if death occurs from other causes during the clinical trial then

(i) if the cause of death is unrelated to breast cancer then treat these deaths as censored observation,

(ii) if the cause of death is related to breast cancer then treat these as death time.

My comment is that if the number of deaths is small then it doesn't really matter, whether you use (i) or (ii), but if the number of deaths from other causes is substantial then very misleading results can be obtained because of the non-identifiability problems of competing risks.

5

Dr. Haybittle

I cannot really agree with your suggestion that deaths from intercur-

rent disease should be treated as withdrawals at their time of death. In my view, the inaccuracies of death certification and our often very limited knowledge of the patient's condition in the period immediately preceeding death make it almost impossible for us to be sure that death was not either due to cancer at the site treated or due to some cause related to treatment. I would prefer to see all such deaths included as deaths when comparing survival curves in a clinical trial.

Answer to Dr. Tsiatis and to Dr. Haybittle

6

Dr. Palmer

The suggestion that deaths from intercurrent disease should be treated as censored observations is valid providing that there is enough reliable information to be able to distinguish between deaths from the treated cancer and deaths from other causes, including other cancers. I do agree with Dr. Haybittle that death certificates are often inaccurate and we would not rely solely on these. In Manchester patients are kept on very close clinical follow-up and so we believe it is possible to identify some deaths as being completely unrelated to the breast cancer or its treatment and the statistical analysis will be slightly more sensitive if these are treated as censorings.

I agree, however, with the remark made by Dr. Tsiatis that if the number of intercurrent deaths is small, then it really does not matter whether they are treated as censored or exact survival times. So far as his remark concerning non-identifiability is concerned, I understand this to mean that a survival curve calculated by treating intercurrent deaths as censored observations is not the curve we would observe in the absence of all causes of death other than breast cancer, i.e. if all other diseases could be cured, and to this extent the curve is estimating a rather nebulous concept. However, this does not matter since the purpose of a trial is to make inferences about the relative merits of two different types of treatment and these inferences are still valid, providing the mechanism of intercurrent death acts independently of treatment group membership.

Comment to Dr. Palmer's and to Dr. Ribeiro's paper

7

Dr. Prentice

I would like to make a small contribution to the discussion on strati-

fication. In trials that I am familiar with a sensible degree of stratification can be incorporated without any noteworthy complication in the conduct of the study. In fact, this seems like one of the few areas in clinical studies in which some control can be readily exercised. The benefits of such control have been recognized by statisticians for many years. Though these are somewhat less evident in a very large trial some efficiency gain in treatment comparisons can be anticipated on the basis of stratification at the design stage, and perhaps substantial efficiency gains will arise in the study of the interaction of treatment effects with stratification factors. An additional point is that simple survival curve presentations by treatment group are more readily interpreted if important prognostic factors have been balanced by stratification. Of course it should be pointed out that stratification on too many factors becomes rather meaningless and may virtually be equivalent to no stratification whatever.

The King's/Cambridge Trial

Michael Baum

King's College Hospital Medical School
(University of London)

Introduction

In 1842 Sir James Syme[1], Professor of Surgery in Edinburgh stated; "the results of operations for carcinoma when the glands are affected is almost always unsatisfactory however perfectly they may have seemed to be taken away. The reason of this probably is that the glands do not participate in the disease unless the system is strongly disposed to it". It has taken us well over a hundred years to rediscover this fundamental truth concerning the biological nature of carcinoma of the breast.

Late in the last century Halsted described his radical mastectomy, the operation was designed on the basis of the pathological teachings of those times popularised by Sampson Handley a surgeon working at the Middlesex Hospital[2]. It was believed that the cancer started as a single focus within the breast and then spread centrifugally as continuous columns of cancer cells along the lymphatics until the tumour became arrested at the regional lymph nodes. These 'glands' were thought to act as filter traps controlling further spread of the cancer until they became exhausted, following which the tumour spread again through the efferent lymphatics reaching the vital organs such as the liver and brain in direct continuity with the primary tumour. It was also believed the involvement of the bone did not occur unless there was a skin nodule superficial to the site of the metastasis thus maintaining the integrity of the webb of cancer spreading out from its central focus. It therefore seemed perfectly reasonable at the time that the wider the field of surgery and the greater number of lymph nodes incorporated in the surgical specimen the greater number of cures. However there was no prima-facie evidence to support this point of view and the inconsistencies in this particular biological model have already been well documented and reviewed on many occasions[3,4]. It is useful to remember some of the advances in our understanding of the biology of cancer in the last 50 or 60 years which in particular make the Halsted/Handley hypothesis untenable. Firstly, it is recognised that cancer can gain access to the blood stream directly by-

passing the lymphatics[5]. Secondly, cancer cells may indeed embolize along lymphatic channels but can either bypass or traverse the lymph nodes which are ineffective as filters[6], and finally the lymph nodes far from being inert must be considered actively hostile to the implantation of cancer cells as they are rich in lymphocytes and histiocytes which in vitro are known to possess tumouricidal properties[7]. Taking account of the new understanding of the biology of cancer, Devitt in 1965 described an alternative hypothesis which he succintly summarised as follows "Axillary lymph node metastases are an expression of a bad prognosis rather than determinant"[8]. Further development of this idea could suggest that a woman with primary carcinoma of the breast and negative axillary nodes is not necessarily exhibiting a chronologically early cancer but one that is biologically favourable, with the regional lymph nodes successfully destroying any cancer cells that may have arrived within their substance and perhaps playing a role in the maintenance of systemic immunity. Interference with these uninvolved nodes either by surgery or radiotherapy, would in theory be to the detriment of the patient[9]. Whereas considering the woman with extensively involved axillary nodes, we know that in the majority of cases "however perfectly the lymph nodes may have been taken away" the disease will inevitably relapse leading to the patient's death[10]. The involvement of the axillary nodes in these cases may therefore represent an exhaustion of the tumour host balance symptomatic of a systemic disease. However it could still be considered that radical surgery in such cases may have a role to play. The putative host defense mechanisms are known to be weak and in experimental models can only cope with between 10^7 and 10^8 cancer cells. Therefore local radical surgery or radiotherapy, destroying as much as possible of the tumour burden of the host, may yet tip the balance in favour of host defense mechanism which perhaps could cope with residual micrometastases within the vital organs.

The above summarizes the debate that was actively raging in the United Kingdom ten years ago that lead to a meeting being called in Cambridge, by groups of clinicians working at Addenbrookes Hospital, Cambridge and King's College Hospital, London, for the purpose of launching a trial that once and for all would resolve this particular dispute.

Problems of designing protocol

Right from the outset there were problems in defining the protocol for a trial that would answer the questions which can perhaps best be summarized as follows:-

a) Does the pathologically involved regional lymph nodes act as a site for tertiary spread of the cancer, whereby leaving such nodes untreated reduces the survival prospects for the individual patient?

b) Does radical treatment either by surgery or radiotherapy so reduce the natural defense mechanisms of the patient as to facilitate the outgrowth of micrometastases disseminated at the time of diagnosis?

On the basis of these questions a radical and a conservative method of treatment for primary carcinoma of the breast that was acceptable to the majority of the clinicians in the United Kingdom had to be defined. Fortunately at that time a 'popularity poll' had been published in the British Journal of Surgery which demonstrated that the largest percentage of surgeons in the United Kingdom were practising the McWhirter regimen for early breast cancer with simple mastectomy followed by radiotherapy to the chest wall and the regional lymph node fields[11]. This then could be considered the standard 'radical approach' to the management of primary carcinoma of the breast in vogue at the time. A conservative approach therefore would be simple mastectomy alone with no surgical interference with the axillary nodes and radiotherapy delayed until there was local recurrence on the chest wall or obvious progression of the axillary lymph nodes. It was easy enough to describe the operation of simple mastectomy but considerable discussion was necessary to decide the radiotherapy protocols and finally it was elected to use a dose schedule that was again widely accepted within the United Kingdom. This is summarised in Appendix 1 of this paper which outlines the detailed protocol of the trial. Finally it was agreed that a trial should be launched which if nothing else had the virtue of simplicity of design. All patients presenting with clinically Stage I or clinically Stage II carcinoma of the breast should be randomised into a 'radical group' or a 'watch policy' group. The radical group would have simple mastectomy with immediate radiotherapy and the conservative or watch policy group would have simple mastectomy alone with careful observation of the axilla and radiotherapy delayed until there was obvious local progression or recurrence of the disease.

The numbers game

The King's/Cambridge Trial later to be christened the Cancer Research Campaign Trial (thanks to the generous financial support of the C.R.C.), is of great historical importance because for the first time a trial was designed specifically to minimise the chance of a type 2 error[12]. It was postulated from the outset that if any differences in survival

between the two treatment programmes really did exist the order of the difference would be small, say 10%. Furthermore if a difference of 10% or more did exist, then it would be of great clinical importance as well as biological interest to detect this difference. An extrapolation of such a finding across the national mortality statistics for carcinoma of the breast would imply the salvage of 1,000 women's lives each year in the United Kingdom. Therefore, based on the statistical tables described by Boag and his colleagues[13], it was decided from the outset to accumulate 2,000 patients, i.e., 1,000 in each sub-group, in order to stand a more than 90% chance of detecting a less than 10% difference in survival between the two groups at a level of significance that would be statistically acceptable.

Problems of Recruitment

Having decided to recruit 2,000 patients into a trial it was quite obvious that as many centres as possible would have to be encouraged to join the protocol. At the same time, in order that the enthusiasm and interest of the participants would not wane, it was also decided to attempt to collect this number within three years. A small team of surgeons from the Department of Surgery at King's College Hospital were therefore sent on an itinerary around the United Kingdom on a recruitment drive. In addition, letters were published in The Lancet and the British Medical Journal, advertising the trial. After about six months intense activity, 80 centres were encouraged to join not only in the United Kingdom but also as far afield as Northern Europe, New Zealand and the far West of Canada. A list of participants appears on Appendix 2.

The first patients were entered into the trial in June 1970. Recruitment continued at a very satisfactory rate in excess of the predicted requirements so that over 2,800 patients were submitted for inclusion in the Study. The successful recruitment of such large numbers in such a short time can probably be ascribed to three factors.

a) The question asked was uppermost in most clinician's minds.
b) The design of the protocol was extremely simple.
c) Many clinicians working in hospitals remote from the major teaching centres were excited by the prospect of involvement in an international collaborative randomised trial.

Thus, for the first time an attempt was made to tap the enormous well of goodwill that existed amongst surgeons practising in the National Health Service.

Problems

It cannot be denied, that enormous problems were encountered in the early days of this study in handling the vast amount of data in what had rapidly become the largest clinical trial ever undertaken for the treatment of cancer. The Study was being run by a group fo junior clinicians completely inexperienced in clinical trial methodology. It soon became appreciated that the Trial Secretariat was hopelessly understaffed, but furtunately the Cancer Research Campaign generously supported additional staff to cope with the workload. Problems were encountered in writing programmes for the computer, for storing and analysis of the data. After several attempts it was necessary to transfer all the information to a new computer centre at St. Mary's Hospital, Paddington. In retrospect therefore, involvement of a competent computer staff in the design of the protocol, proformas and data handling systems, right from the start, would have avoided a lot of the early problems. In the early days of the study there were en enormous number of protocol violations, which reflected the inexperience of the participating clinicians in entering the patients into such trials, but by far and away the biggest problem was encountered in the randomization procedures. It was wrongly decided at the outset that each surgeon should keep a bundle of brown paper envelopes, labelled in sequence, containing the random allocation cards. This system failed on many occasions, and in addition there was reason to believe that individual surgeons were actually abusing the system and breaking the code before deciding to enter the patient into the study. Fortunately, a central allocation record had been kept and where patients from an individual centre were entered out of sequence, the Trial Coordinator paid a visit to the individual centre to confirm that such discrepancies were or were not an active abuse of the system of peripheral randomization. As a result of this problem several centres had to be excluded in toto from the analysis of the study, thus losing in excess of 500 patients from analysis, but as each centre had its own random series such exclusions do not invalidate the overall analysis of the trial.

Results

After exclusion of ineligible patients through protocol violations or randomization errors, there were 2,224 evaluable patients left for analysis. All patients have been followed up now from between two and seven years. A full report with a five year actuarial analysis was published in the British Medical Journal in 1976 and the position over

the last two years had not materially altered[14].

Local Recurrence

There is a highly significant difference in favour of the radiotherapy group in the incidence of local recurrence, with approximately 30% of patients in the watch policy group developing local recurrence compared with about 5% in the irradiated group. However, the local recurrences are predominantly in the untreated axillary nodes which have progressed on observation and are probably better referred to as persistance of untreated disease rather than true recurrence. Although data on this point is incomplete the early impression is that the majority of patients having delayed radiotherapy in the watch policy group achieve local control of the disease. However, in the final analysis the incidence of uncontrolled local recurrence at the time of death in the two groups will be the most important fact upon which to base therapeutic judgement in the future and this information is not yet available.

Survival

The paper published in the British Medical Journal demonstrated actuarial survival curves for the two treatment groups that were virtually superimposed and with the period of observation extended two more years there is no further evidence to suggest that one therapeutic strategy is superior to another as far as survival is concerned. Subdividing patients into those that were clinically Stage I and clinically Stage II on entry into the trial, there were 1,688 evaluable patients in the former category, again with no significant difference as far as survival is concerned, and 536 patients with clinically involved nodes, again demonstrating no significant difference in survival. Full statistical details with up-to-data actuarial survival curves are shortly to be published and are not particularly relevant in the context of this meeting.

Biological Fall Out

When conducting a clinical trial comparing two treatments, it is an oversimplification just to think of the exercise as enabling you to choose the 'best buy'. As well exemplified in the King's/Cambridge Trial, treatment A versus treatment B was, infact, testing hypothesis A against hypothesis B. Inevitably, therefore, there will be biological fall-out from such a study, the information thereafter being built

into a new hypothesis to describe the behaviour of the disease which within time will generate a new treatment C. This aspect of scientific philosophy has been expanded upon in previous publications by the author[15,16]. However, it is worth mentioning one interesting and unexpected finding which has appeared as a result of this study. In the early years after the establishment of the trial it rapidly became apparent that the untreated axillary nodes in the watch policy group of patients, far from all rapidly progressing and demanding treatment, behaved in an unexpected manner. Edwards, Baum and Magarey[17] described the experience of this group of patients in the first twelve months of follow-up. Forty clinically Stage II patients were described and 75% of these no longer demonstrated palpable axillary nodes after three months, whereas of the 120 clinically Stage I patients (of whom we know at least 30% probably have pathologically involved nodes[18]) only 10% demonstrated clinical progression of the disease over the first twelve months after mastectomy. This trend has continued throughout the follow-up of the trial and a more detailed analysis by Baum and Coyle of a sub-group of these patients followed for up to four years has highlighted an interesting clinical phenomenon[19]. Those patients who demonstrated a clinical regression of palpable nodes seem to have a favourable prognosis, suggesting that resolution of a reactive hyperplasia within the axillary nodes, whereas those patients whose nodes have progressed on observation almost all develop distant metastases within a median period of eight months. It must be emphasized however, that these are merely clinical observations and it has never been claimed that pathologically involved nodes can undergo spontaneous regression.

Discussion

Based on the interim results of this trial, one is forced to conclude that on the one hand untreated axillary nodes probably do not act as a source of tertiary spread of disease, whereas on the other hand regional radiotherapy, although known to produce a systemic immunosuppression[9], does not significantly affect the outcome as far as survival. This suggests that the radiologically induced lymphocytopenia is of no clinical significance. Furthermore, based on the behaviour of the untreated lymph nodes, we have further evidence to support the idea that the natural history of the disease within the axillary lymph nodes merely reflects the behaviour of the disease within the viscera or skeleton, and on this basis axillary lymph node metastases could be considered as an accessible and easily recognised manifestation of

systemic disease. All these findings and their interpretation are supported by the experience of the NSABP clinical trial, Protocol 4, published by Fisher and his colleagues in 1975[20]. It is still conceivable that a 15 year follow-up may, infact, demonstrate differences in long term survival related to the neglect in treating accessible disease within the axilla. Meanwhile because of the large numbers within this study and therefore its statistical power, we can state with some degree of confidence that if a real difference exists at five years it is probably less than 7%.

Conclusion

It was not the intention of this talk to instruct my colleagues how to treat early carcinoma of the breast. On the basis of the results described the clinician still is faced with the dilemma, should he spare the majority of women radiotherapy who are unlikely to develop local recurrence of the disease, or should he irradiate all women after mastectomy in order to protect, say, 30% of patients from an awareness that treatment has failed, assuming that both groups are dying off at the same rate? This decision is entirely a value judgement. If nothing else, the data generated by this clinical trial has enabled us to add weight to the hypothesis suggested by Fisher[6] (and over 20 years earlier by MacDonald[21] and McKinnon[22]), that the outcome of treatment of early carcinoma of the breast is predetermined by the extent of micrometastases at the time of diagnosis and not influenced by the extent of local therapy. This alternative hypothesis is now being vigorously tested throughout the world by trials of adjuvant systemic therapy.

References

1. SYME, J. (1842). Principles of Surgery. London, Bailiere.

2. HALSTED, W.S. (1907). Annals of Surgery, 46, 1.

3. CRILE, G. Jr. (1967). A biological consideration of treatment of breast cancer. Fort Lauderdale,Florida,Thomas

4. BAUM, M., (1977). The curability of breast cancer. In Breast Cancer Management - Early and Late. Ed. B.A. Stoll, London, William Heinemann

5. GRIFFITHS, J.D. and SALSBURY, A.J. (1965). Circulating cancer cells Thomas, Springfield, Illinois.

6. FISHER, B. (1970). Current problems in Surgery, Chicago, Year Book Publishers.

7. HELLSTROM, I., HELLSTROM, K.E., SJOGREN, H.O. and WARNER, G.A. (1971) International Journal of Cancer, 7,1.

8. DEVITT, J.E. (1965). Canadian Medical Association Journal,93,289.

9. STJERNSWÄRD,J., JONDAL, M. VANKY, F. et al. (1972) Lancet 1, 1352

10. FISHER, B., SLACK, N.H., and BROSS, I.D.J. (1969). Cancer, 24, 1071.

11. FORREST, A.P.M. (1969). British Journal of Surgery 56, 782.

12. FRIEMAN, J.A. CHALMERS, T.C., SMITH, H., and KUEBLER, R.R. (1978) The New England Journal of Medicine 299, 690.

13. BOAG, J.W., HAYBITTLE, J.L., FOWLER, J.F. and EMERY, E.W. (1971). British Journal of Radiology 44, 122.

14. Cancer Research Campaign Working Party (1976). British Medical Journal, 1, 1035.

15. BAUM, M. (1975) Bulletin du Cancer, 62, 391.

16. BAUM, M. (1979) Current controversy in the management of early carcinoma of the breast. In Surgical Review, Editors Lumley & Craven, Pitman Medical, Tunbridge Wells.

17. EDWARDS, M.H., BAUM, M. and Magarey, C.J. (1972). British Journal of Surgery, 59, 776.

18. Yorkshire Breast Cancer Group (1977) British Medical Journal 2, 1196.

19. BAUM, M. and COYLE, P.J. (1977) Bulletin du Cancer 64, 603.

20. FISHER, B., MONTAGUE, E., REDMOND, C. et al. (1977) Cancer 39, 2827.

21. MacDONALD, I. (1951). Surgery, Gynaecology & Obstetrics,92, 443.

22. McKINNON, N.E. (1954). Lancet 1, 251.

APPENDIX 1

BREAST TRIAL PROTOCOL

1. Introduction

Carcinoma of the breast is the most common form of cancer in women, yet the most effective method of treatment is not known. It has been assumed that radical treatment of the regional lymph nodes by surgery or radio-therapy is necessary, but no evidence has been produced to show that such treatment is of benefit. Harm may have been done.

It is proposed to carry out a prospective, controlled clinical trial to determine whether the routine treatment of the regional lymph nodes is of benefit to patients with apparently early carcinoma of the breast. It may be possible in the future to spare many such patients the discomforts of treatment which may not only be unnecessary, but positively harmful.

2. General Principles

(i) Aims:

(a) To determine whether treatment of the regional lymph nodes in patients with clinically early carcinoma of the breast
- affects the survival of the patients.
- affects the site or time of recurrence of the tumour.
- affects the morbidity of the patients.

(b) To determine if there are any clinical or pathological features which may influence the choice of treatment of these patients in the future.

(ii) Types of Treatment

(a) **Irradiated Group:**
Simple mastectomy and routine postoperative regional radiotherapy.

(b) **Watch Policy Group:**
Simple mastectomy alone, aiming to avoid what may be unnecessary treatment of the lymph nodes.

(iii) Follow-up Policy

The aim of follow-up for all patients is to detect any actively growing tumour recurrence or metastasis, and then to treat it by whatever means are appropriate in the particular patient. There is no intention of allowing lymph-node or other metastases to progress to the extent that symptoms may arise.

When may the axilla be treated?

The following criteria should be regarded as evidence of active growth of tumour in the axillary lymph nodes and will be defined as 'Local recurrence in the Ipsilateral Axilla".

(a) Clear evidence of progressive enlargement of a lymph node.
(b) Development of fixation of a lymph node.
(c) Appearance of symptoms referable to the axilla.

Treatment of the axilla should not be carried out until one of these criteria is fulfilled. This applies to patients in either treatment group, particularly to those in the Watch Policy Group, whose axillary nodes may be palpable after simple mastectomy. It is unlikely that these axillae will require treatment within three months of mastectomy.

Any appropriate mode of treatment may be used.

3. Clinical Criteria for Inclusion of Patients

All patients who satisfy the following criteria should be included in the trial.

Patient

- less than 70 years of age.
- a woman with apparently early carcinoma of the breast, who is willing and fit for treatment by mastectomy and radiotherapy.
- no previous treatment for malignant disease.

Tumour

- 5 cm. or less in its greatest dimension.
- incomplete skin fixation **or** no skin fixation.
- no fixation to chest wall or pectoral muscles.

Nipple
— may be normal
or retracted
or involved by Paget's disease, which may extend beyond the nipple.

Lymph Nodes
— ipsilateral axillary lymph nodes, if palpable, should be mobile.

Metastases
— no evidence of metastasis on clinical examination or radiography of chest and axial skeleton.

These criteria are identical with those of the International Union Against Cancer Classifications (Stage I and II; **T1** and **T2; N0 and N1; M0**).

4. **Selection of Individual Patients**

Ideally, the Surgeon will include in the Trial all patients with palpable or with impalpable axillary lymph nodes. (i.e. all Stage I and II cases).

If he particularly wishes, he may exclude all patients with palpable axillary nodes (Stage II). or all patients without palpable axillary nodes (Stage I).

5. **Random Allocation of Patients**

Each patient will be allocated to the "Irradiated Group" or to the "Watch Policy Group" **after** Standard Simple Mastectomy has been performed. Thus, it is only after operation and after the diagnosis of carcinoma has been confirmed, that the patient finally enters the Trial. It is important to avoid allowing prior knowledge of subsequent treatment to influence the operative technique.

A sequence of numbered, sealed Allocation Envelopes will be supplied to each hospital.

One envelope will then be opened for each patient entering the Trial. Inside the envelope will be found instructions for the random allocation of the patient to one or other treatment group. The instructions will be arranged so that approximately equal numbers of patients at each hospital join the Watch Policy Group and the Irradiated Group.

6. **Exclusion of Patients**

All Surgeons have a right to exclude any patient from the Trial, but must do so before a Random Allocation Envelope has been opened. This right will need to be exercised only under exceptional circumstances and records should be made of the patient's name and of the reason for her exclusion.

Once a patient has been allocatd to one Treatment Group, she **must** remain in the Trial and in that particular Treatment Group. This applies even if the patient, for some reason, does not follow the planned treatment regime, e.g. refusal by the patient to have radiotherapy.

7. **Timetable for Admission of Patients to the Trial**

Step 1 The Patient with a Breast Lump, which may be a carcinoma, is admitted to Hospital.
Step 2 A Proforma is placed securely in the Patient's Case Sheet.
Step 3 A blue sticker is attached to the Front Cover of the Case Sheet.
Step 4 Admission Sheets One and Two are filled in **before** operation.
Step 5 A Biopsy of the Breast Lump is performed if necessary.
Step 6 A Standard Simple Mastectomy is performed.
Step 7 The Surgeon commences to fill in the Operation Sheet.
Step 8 The Patient's Name and Hospital Number are written on the Pathology Sheet.
Step 9 The Pathology Sheet and **all** the Operation Specimens are sent to the Hospital Pathologist. (He eventually fills in his part of the Pathology Sheets, and sends it with Blocks or Slides of the specimens, and a copy of his Pathology report, to the Breast Trial Centre.)
Step 10 The Surgeon obtains the Histological Diagnosis of the Breast Lump.
Step 11 **If carcinoma is not present,** the Surgeon excludes the Patient from further study. The Proforma is discarded and no Allocation Envelope is opened.
If carcinoma is present, the next numbered Allocation Envelope is opened.
Step 12 The Patient's Breast Trial Code Number is written on the two Allocation Cards which are found inside the Envelope.

Step 13 The Patient's allocation is recorded on the Operation Sheet, which is thereby completely filled in.

Step 14 **If the Patient is in the Watch Policy Group,** the Surgeon passes directly to Step 16.
If the Patient is in the Irradiated Group, she is referred to the Radiotherapy Unit.

Step 15 **If the Patient is in the Irradiated Group,** the Radiotherapy Sheet and one Allocation Card are sent to the Radiotherapy Consultant concerned.
(The Patient eventually receives radiotherapy according to the regime described in this Protocol. The Radiotheraphy Sheet of the Proforma is filled in at the Radiotherapy Unit, and is then sent to the Breast Trial Centre.)

Step 16 **If the Patient is allocated to either Treatment Group,** the 3 Month Follow Up appointment is arranged.

Step 17 The Surgeon sends to the Breast Trial Centre :—
One Allocation Card
The Copy of Admission Sheet One
Admission Sheet Two
The Operation Sheet.

Step 18—23 The Patient will be seen in a Follow Up Clinic at the times indicated on the Proforma. A Follow Up Sheet is filled in each time and is sent to the Breast Trial Centre. Centre.

The Patient may be seen more frequently, if necessary, but no Follow Up Sheet should be filled in on these additional visits.

When a Patient is referred to another hospital for Follow Up, her Proforma should be placed in the Case Sheet of the Follow Up hospital. A blue sticker is placed on the front cover of that Case Sheet.

If the Patient fails to keep a Follow Up appointment, the reason should be determined and arrangements be made for her to attend the Clinic. It is most important that the Patient should be followed very closely. To assist the Clinicians, the Computer will determine which Patients' Follow Up Sheets have not reached the Breast Trial Centre at the expected time. This information will be passed on to Clinicians concerned.

Step 24—25 If the Follow Up Sheet records the development of a Recurrence or a New Tumour in the Patient, a request for further details will be sent to the Clinicians concerned after a period of 6 months. In this way, a record of the Patient's response to any treatment can be made.

Step 26—27 If the Patient dies, the Record of Death in the Proforma is filled in and is returned to the Breast Trial Centre.

8. **Standard Simple Mastectomy**

Aims To remove all the breast
To avoid interfering with the lymph nodes as far as it is possible.

Procedure

The skin incision is planned to include the nipple and areola, and the skin overlying the tumour and any biopsy incision.

Whether or not the skin flaps have been made deliberately thin, is recorded.

The breast with its axillary tail is removed.

Whether or not the deep fascia has been removed, is recorded.

While one or more lymph nodes may be found incidentally by the Pathologist in the tail of the amputated breast, deliberate excision of lymph nodes from the axilla should be avoided by the Surgeon. Any enlarged nodes in the axilla should be left undisturbed.

Internal mammary nodes should not be removed.

N.B. In order to standardise the initial treatment of patients in the Trial, the following **prophylactic** methods should **not** be used :—
Endocrine operation
Hormone therapy
Cytotoxic drugs
Immunotherapy

When a Recurrence appears and requires treatment, naturally such measures may be undertaken.

9. Pathology

The Hospital Pathologist will receive the operation specimens together with the Pathology Sheet from the Proforma.

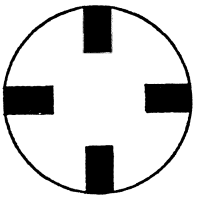
It is intended to investigate the histological features of a tumour/host response in the Breast Trial patients. In order to obtain uniformity, two pathologists (at King's College Hospital and Addenbrooke's Hospital) will examine sections from every tumour and will be responsible for completing the Microscopic Appearances Section of the Pathology Sheet.

All that is requested of the Hospital Pathologist is :—

(i) Completion of the column on the Pathology Sheet of the Proforma entitled "Macroscopic Appearance".

(ii) Loan of :

(a) 4 blocks, or 12 unstained spare slides from each of four blocks, of the primary neoplasm. It is suggested that the tumour be sampled in a cruciate way to produce four blocks thus :—



(b) the blocks, or 12 unstained spare slides from each block, of lymph nodes that have been found. Any lymph nodes that can be found in the specimen should be included.

(c) blocks or spare slides of nipple skin.

Any slides or blocks will be returned by registered post.

(iii) A copy of the Hospital Pathologist's report of the macroscopic and microscopic appearances of the breast.

10. Standard Radiotherapy

In this trial of treatment of early breast cancer it is important that the skin flaps and all the lymphatic drainage areas are irradiated. It is suggested that the skin flaps, axilla, supraclavicular and internal mamary areas should be treated with ortho- or supervoltage radiotherapy.

The minimum deep tissue dosage on the chest wall, the doses at the mid-point of the axilla, and at the estimated depth of the supraclavicular and internal mammary glands should lie within the values given below in **rads.**

	(18 days) 3 weeks	(25 days) 4 weeks	(32 days) 5 weeks	(39 days) 6 weeks
3 fractions per week	2,850—3,150	3,200—3,500	3,450—3,850	3,700—4,050
5 fractions per week	3,250—3,600	3,650—4,000	3,950—4,350	4,200—4,600

Multiply by 1.1 for supervoltage

It is hoped that the above range of doses and times will cover most Radiotherapists' usual post-mastectomy treatment programmes. Small differences in dosage and time will inevitably occur and need not matter, providing they are recorded. If any radiotherapist uses a technique which is widely different from the suggested schemes, he should get in touch with the Breast Trial Organisers at King's College Hospital.

It is intended that the Radiotherapist should assess the degree of early skin reaction I month and 2 months after the course of radiation has ended. Late Radiation Telangiectasis is assessed in the Follow-Up Clinic.

11. Follow-up Duration

Over 600 patients are required in this Trial to satisfy statistical criteria.

It is expected that over 1,000 will, in fact, be collected within a period of 2 years.

The follow-up period will extend over the following 10 years or more.

Data should be forwarded to the Breast Trial Centre 3-monthly for 18 months, 6-monthly for 1 year, and yearly for the following 8 years. It should be noted, however, that clinicians are free to review patients more frequently if they wish and as the clinical condition demands.

12. Analysis

Recurrence, morbidity and survival rates will be calculated every six months using the computer at King's College Hospital and the results will be communicated to all participators.

Similarly, details of any pathological or clinical features which correlate with improved or impaired results in treatment will be circulated to those taking part.

It is expected that progress reports will be published annually in the medical press.

Appendix 2

Participants in the study were:

Mr.J. Kyle, Mr.J. Nelson Norman, Dr.J. Carr, the late Dr. E. Ridley, Aberdeen, Royal Infirmary; Professor J.S. Mitchell, Professor R.Y. Calne, Professor G.A. Gresham, Mr. A. Conway, the late Mr. B. Truscott, Mr. W.G. Everett, Mr. A. Smellie, Dr. T. Wheeler, Dr. G. Bratherton, Dr. E. Kingsley Pillers, Dr. G. Rao, Addenbrooke's Hospital, Cambridge; Mr. S.T. McCollum, Professor W. McCaughey, Adelaide Hospital, Dublin; Mr. H.M. Bennett, Dr. J.S. Elwood, Altnagelvin Hospital, Northern Ireland; Mr. W.A. Gallagher, Mr. O.H.A. Mitchell, Ards Hospital, Northern Ireland; Nr. D.M. Bell, Mr. R.C. Curry, Mr. W.A.Hanna, Dr. J. Morrison, Belfast City Hospital; Dr. F. Kelly, Belvidere Hospital, Glasgow; Mr. H.M. Jamison, Mr, J.H. Wrigley, Dr. J.M. Robertson, Bishop Auckland Hospital, Co. Durham; Dr. T. Castberg, Dr. C. Johansen, Dr. Hammer-Jacobsen, Dr. J. Rygard, Bispebjerg Hospital, Denmark; The late Professor A.G. Riddell, Professor J. Peacock, Mr. K. Hobbs, Dr. R. Tudway, Dr. H. Eckert, Dr. B. T. Hale, Professor M. Epstein, Bristol Royal Infirmary; Mr. J. Dawson, Dr. D. Brinkley, Dr. M. Millard, Bromley Hospital; Dr. D. O'Connell, Charing Cross Hospital; Mr. J. Fairgrieve, Mr. R. Harvey, Dr. F.A. Hanna, Dr. A. Nicol, Dr. J. Pitt-Evans, Cheltenham General Hospital; Mr. W.M. Gray, Mr. E.B.Z. Masterman, Mr. R. Blamey, Dr. R.E. Cotton, Dr. J.S. Jones, Dr. P.G. Smith, City Hospital, Nottingham; Dr. T.W. Backhouse, Dr. K. Sicher, Coventry and Warwickshire Hospital; Mr. F.M. Hanna, Mr. W.A. Tucker, Dr. K.S. Holmes, Dr. A.N. Blades, Dr. T.V. Cooper, Dorset County Hospital; Mr. J.S. Boyd, Downe Hospital, Northern Ireland; Mr. G.H. Dunstone, Mr. R. Petticrew, Dr. J. Ennis, Dryburn Hospital, Co. Durham; Professor J.G. Murray, Dr. D. Brinkley, Dr. C. Elston, Dulwich Hospital, London; Dr. C. Swanson, Dundee Roayl Infirmary; Professor A. Clarke, Mr. M.E. Dawson, Mr. J.H. Heslop, Professor W. Macbeth, Mr. R.F.X. Norhona, Mr. V.T. Pearse, Mr. M. Shackleton, Professor J. Blennerhassett, Dr. N. Fithgerald, Dr. D. Perry, Dunedin Hospital, New Zealand; Mr. M.F. Hunt, Mr. R.N. Jones, Mr. D.M. Millar, Mr. N. Orr, Dr. D. Gamble, Dr. Rhys-Lewis, Dr. G.S. Anderson, Essex County Hospital; Mr. M. J. Solan, Dr. R.W. Ainsworth, Farnham Hospital; Mr. J. Anderson, Professor R. Goudie, Glasgow Royal Infirmary; Dr. L.H. Walter, Hammersmith Hospital, London; Mr. G.H.D. McNaught, Dr. R.T. Cooke, Dr. H.McTaggart, Dr. R.S. Bundi, Hartlepool General Hospital; Mr. J.S. Darling, Dr. J. Dean, Huntingdon County Hospital; Mr. K.C.D. Gordon, Dr. K.J. James, Dr. A. Lintott, Dr. T. Shaw, Ipswich and East Suffolk Hospital; Professor P.G. Collins, Professor D. Doyle, Jervis Street Hospital, Dublin; Professor G. Martz,

Professor H. Schwarz, Dr. M. Landott, Kantonsspital, Limmattalspital, Triemli, Zürich; Mr. J.H.C. Phillips, Dr. P.S. Andrew, Kettering and District Hospital; Professor J.G. Murray, Mr. J. Dawson, Mr. H. Berry, Dr. D. Brinkley, Dr. C. Elston, King's College Hospital, London; Mr. D. W. Bain, Mr. J.W. Blaxland, Mr. R.J. Luck, Dr. K. Dempster, King Edward VII Hospital, Windsor; Dr. A. MacFarlane, King's Mill Hospital; Mr. M. V. Sheehan, Dr. B.P. O'Flynn, Lady of Lourdes, Drogheda; the late Mr. Vause-Greig, Mr. G.I. Young, Lagan Valley Hospital, Northern Ireland; Dr. H. Hope-Stone, London Hospital; Mr. D.H.C. Harland, Mr. R.V. Fiddian, Mr. W. Mee, Dr. J. Bradley-Watson, Luton and Dunstable Hospital; Mr. M.J. Ball, Mr. A.McEwen Smith, Mansfield and District Hospital; Professor W.P. Hederman, Professor S.J. Heffernan, Professor E.O'Malley, Professor M. Hickey, Mater Hospital, Dublin; Mr. W.D. Mackay, Mr. J.S. Kinnear, Dr. G.H. Smith, Maryfield Hospital, Dundee; Mr. R. Stinson, Massereene Hospital, Northern Ireland; Mr. J. Bradbeer, Dr. D.O'Brien, Mayday Hospital, Croydon; Professor N. Bleehen, Middlesex Hospital; Mr. W.A. Brennen, Mid-Ulster Hospital, Northern Ireland; Dr. S. Dische, Mount Vernon Hospital; Mr. J.G. Kinley, Mr. M. Laird, Moyle Hospital, Northern Ireland; Mr. R. Lavelle, Navan Hospital, Eire; Mr. I.R. Isaac, Mr. J.S.McConnachie, Mr. R.R. Rintoul, Dr. J.Dearnaley, Nevill Hall Hospital, Abergavenny; Mr. A.H. Petty, Mr. R. Finney, Mr. W.M. Ross, Dr. B.J. Smith, Professor B. Tomlinson, Newcastle General Hospital; Mr. R.E. Tagart, Newmarket General Hospital; Dr. G.A. Edelstyn, Dr. A. R. Lyons, Dr. G. Lynch, Dr. D. Burrows, Northern Ireland Radiotherapy Centre; Mr. K. Cronin, Mr. D. Lambley, Dr. H. Cole, Dr. B. Jolles, Northampton General Hospital; Mr. S.G. Thomson, North Cambridgeshire Hospital; Mr. R.A. Payne, Dr. V. Levison , Dr. W. Harrison, North Middlesex Hospital, London; Mr. J.G. Gray, Mr. L.J. Lawson, Mr. E.R. Monypenny, Dr. B. Ockenden, Dr. A McCall, North Staffs Royal Infirmary; Dr. W. Fraser, Nottingham General Hospital; Mr. W.D. Park, Dr. E. Atkinson, Dr. I. Larkin, Oldchurch Hospital, Romford; Mr. T.A. Boxall, Mr. R.M. Walker-Brash, Dr. J. Darby, Dr. K. J. Randall, Orpington Hospital; Dr. M. McEvedy, Pembury Hospital; Mr. D.W. Bracey, Mr. J.R. Thompson, Mr. T. Young, Dr. P.T. Chopping, Dr. F. Fawcett, Peterborough and District Hospital; Mr. R.C. Shepherd, Dr. M. Marlborough, Poole General Hospital; Mr. T.H. Tweedy, Dr. W.K. Cowan, Queen Elizabeth Hospital, Gateshead; Mr. J.C.F. Townsend, Dr. B.S. Jones, Queen Elizabeth II Hospital, Welwyn Garden City; Professor S.F. O'Beirn, Mr. C.Galvin, Mr. B.M. Murphy, Professor J.D. Kennedy, Regional Hospital, Galway; Mr. K. Vowles, Mr. C. Shaldon, Dr. R.A. Caldwell, Dr. R. Hadden, Royal Devon and Exeter Hospital; Mr. R. Yeo, Dr. J.McMurray,

Dr. A.R. Worssam, Royal East Sussex Hospital; Dr. J. Baker, Dr. R.L. Morgan, Royal Marsden Hospital; Mr. J.S. Mousley, Dr. R. Bamforth, Dr. M. Sworn, Royal Hampshire County Hospital; Mr. N. Porter, Dr. J. De Winter, Dr. D. Melcher, Dr. R. Elliott, Royal Sussex County Hospital; Mr. C.P.Sames, Mr. H.T. John, Mr. N. Pizey, Dr. R.L. Bishton, Royal United Hospital, Bath; Mr. R.C. Shepherd, Mr. J.E. Trapnell, Dr. J. Howells, Dr. D. Parish, Dr. A. Rickards, Royal Victoria Hospital, Bournemouth; Mr. G.W. Johnston, Mr. E. Morrison, Mr. S.D. Clarke, Mr. W. Wilson, Professor H.W. Rodgers, Mr. J.S. Irwin, Mr. R.H. Livingston, Mr. J.D.A. Robb, Professor A.D. Roy, Professor D.L. Gardner, Royal Victoria Hospital, Belfast; Professor I.D.A. Johnston, Mr. R. M.R. Taylor, Mr. P.H. Dickinson, Mr. L.B. Fleming, Mr. J.D.T. Jones, Mr. B.McEvedy, Mr. I.F. McNeill, Mr. C.W. Venables, Professor D.N. Walder, Dr. R.G.B. Evans, Mr. W.M. Ross, Dr. C.J. Thurgar, Professor A. Heppleston, Royal Victoria Infirmary, Newcastle; Mr. J.C.F. Townsend, Mr. P. Stringer, Dr. J. Pugh, St. Albans Hospital; Professor G. W. Taylor, Mr. I. McColl, Mr. J.O. Robinson, Mr. H.B. Ross, Mr. W.S. Shand, Dr. A. Jones, Dr. J.J. Lucey, St. Barholomew's Hospital; Mr. B. Wells, Mr. J.M. Edwards, Mr. A. York-Mason, Mr. B. Flannery, Dr. T. Goodier, St. Heliers Hospital, Surrey; Dr. W. White, St.Luke's Hospital, Guildford; Dr. M.J. O'Halloran, St. Luke's Hospital, Dublin; Mr. K. Lloyd-Williams, Mr. W.F.W. Southwood, St. Martin's Hospital, Bath; Dr. T. Bates, St. Thomas's Hospital, London; Mr. P. Shemilt, Dr. R. Pinkerton, Salisbury General Hospital; Dr. C.J. Wright, Saskatoon Hospital, Canada; Mr. J.C.B. Serjeant, Mr. K.G. F. Mackenzie, Dr. I. Porteous, Shotley Bridge Hospital; Professor W.G. Fegan, Sir Patrick Dun's Hospital, Dublin; Mr. J.R. Thomson, Stamford and Rutland Hospital; Mr. J.B. Lowry, Mr. J.T. Ward, South Tyrone Hospital; Mr. J. Fairgrieve, Stroud General Hospital; Mr. H.C. Jones, Dr. K.A. Irvine, Sunderland Royal Infirmary; Mr. G.H. Darke, Mr. A.C. Akehurst, Dr. E. Harries, Taunton and Somerset Hospital; Mr. C.J.H. Logan, Mr. A. McCalister, the Ulster Hospital; Dr. M. Henk, Velindre Hospital, Cardiff; Dr. D. Shine, Wakari Hospital, New Zealand; Mr. G.A. Court, Mr. A. Rhodes, Mr. J.R. Moffat, Mr. D.A.K. Woodward, Dr. J.W. Black, Walsgrave Hospital, Coventry; Mr. A.G. Horsburgh, Dr. C. Pike, Watford General Hospital; Dr. T.K. Morgan, Wessex Radiotherapy Centre; Dr. K. Halnan, Western Infirmary, Glasgow; Mr. W.A.L. Tucker, Weymouth and District Hospital; Mr. L.R. DeJode, Dr. C. Raeburn, Whipps Cross Hospital, London; Mr. P.W. Seargeant, Mr. S.G. Thompson, Dr. D. Eakins, West Norfolk and Kings Lynn Hospital; the late Mr.K.H. Taylor, Mr.P.H. Lord, Dr. C. Paine, Dr. D. Spencer, Wycombe Hospital, Bucks; Mr. T.M.

Williams, Dr. F. Harris, West Suffolk General Hospital; Mr. G.W. Arthur, Worthing Hospital; Dr. A.B. McCarten, Dr. P.E. Burns, Dr. J. Pearson, Dr. P.W. Davey, Dr. J.W. Magregor, Dr. E. Schloss, Dr. J. Stirrat, Dr. D.G. Young, W.W. Cross Cancer Institute.

Comments to Dr. Baum's paper

1

Dr. Kalbfleisch

In view of the reluctance of some clinicians to place advanced stage II patients on trial, I wonder whether some further analysis of the stage II cases might be worthwhile. With envelope randomisation there is always the danger that two patients, who begin at nearly the same time, will be assigned their treatments by the clinician after both randomisations are known. This could easily introduce a bias in the stage II category against the extensive therapy. Would the data be available ans would there be some advantage in investigating whether the lack of difference between treatment is consistent for the various levels of nodal involvement?

Answer to Dr. Kalbfleisch

2

Dr. Baum

Unfortunately staging in the C.R.C. trial was entirely clinical and we have no reliable information on the pathological status of the axilla in these patients. Certainly we know nothing about the various levels of nodal involvement. There can be no doubt, however, that patients with very gross clinical involvement of the axilla were in some cases excluded before randomization. However, the results of our study can only apply to those within the study and a possible interpretation of our results would be that they only apply to clinical node negative patients, and those patients with a minor degree of clinical involvement.

Operable Breast Cancer with Positive Axillary Nodes: The Experience of the Milan Cancer Institute

P. Valagussa, A. Rossi, G. Bonadonna

Istituto Nazionale Tumori, Milan, Italy

The course of breast cancer following potentially curative local-regional modality, i.e. radical mastectomy (RM) plus or minus postoperative irradiation (RT), has been critically re-evaluated during the past few years. The analysis of prospective controlled studies (4,6) has clearly indicated the limits of treatment based only on anatomical principles (5). Furthermore, the histological status of axillary lymph nodes has proved to be the single most useful prognostic factor in women with operable breast cancer. In fact, slightly less than one fourth of women with histologically positive axillary nodes (N+) remained relapse-free at 10 years compared to about three fourths of patients with negative nodes (N-) (4,6).

In the attempt to change the course of operable breast cancer in patients with positive axillary nodes, other approaches were considered: extensive surgery, postoperative radiotherapy, short-term single agent chemotherapy and prophylactic ovarian castration. So far, none of them has really proven to have a substantial impact on relapse-free and especially overall survival (2).

In the past recent years, the advent of cyclical adjuvant chemotherapy (1,3), has proved to have a favourable effect in reducing the early relapses after conventional surgery, especially in premenopausal patients. Moreover, combination chemotherapy with CMF (cyclophosphamide, methotrexate and fluorouracile) has also significantly improved the overall survival (2). However, the results achieved so far were not considered, in general, convincing enough for a wide application of adjuvant chemotherapy in clinical practice and a longer follow-up observation was deemed necessary to prove whether adjuvant CMF was definitely altering the natural course of operable breast cancer at high-risk of relapse (N+).

The aim of this report is to briefly summarize the results of the 4-years analysis of the CMF randomized study. Furthermore, since in our Institute patient selection, types of surgery and follow-up observations were kept uniform throughout the past 15 years, the data of the

controlled CMF study shall also be retrospectively compared with those of two previous series treated with radical mastectomy alone and radical mastectomy followed by postoperative radiotherapy, respectively. The four series are comparable in terms of tumour size (T_2: tumour measuring 2-5 cm in largest diameter), menopausal status and median age. However, in the series treated with mastectomy plus RT it was not always possible to retrospectively assess in all patients the exact number of positive axillary nodes (1-3 vs. > 3). Furthermore, the percent of women subjected to extended radical mastectomy was different in the first two series mainly because the group of patients treated between 1964 and 1967 belong to a prospective controlled trial (Halsted radical mastectomy vs. extended radical mastectomy). However, this difference was previously demonstrated not to influence per se the natural course of operable breast cancer (6). Postoperative RT was delivered with orthovoltage (96%) or Cobalt-60 (4%) units to ipsilateral internal mammary and supraclavicular nodes. The majority of women also received orthovoltage irradiation to the axilla (86%) while in no patient was the chest wall (skin flaps) irradiated. Doses ranged from 4,000 rads (skin-dose) with orthovoltage to 4,000-4,500 rad (target dose) with cobalt-60, in 5 to 6 weeks.

Table 1 shows that the incidence of treatment failure at 3 and 4 years was not significantly affected by various forms of local-regional therapy. On the contrary, the overall treatment failure was significantly reduced only when patients were given prolonged adjuvant systemic chemotherapy.

At 4 years from mastectomy combination chemotherapy also significantly reduced the incidence of local-regional recurrence. From this point of view, the results achieved with CMF (7.3%) are competitive with those obtained with postoperative RT (7.5%). On the contraty, more than 13% of patients treated only with radical mastectomy showed new disease manifestations in local-regional areas. Table 2 shows that there was no statistical difference in the total failure rate at 4 years between pre- and post-menopausal women among the three series treated with local-regional modality. On the contrary, in premenopausal patients RM + CMF significantly reduced all types of treatment failure compared to RM alone ($P = 0.00001$). In postmenopausal women the 4-year actuarial analysis confirms that the therapeutic effects of adjuvant CMF were not significant when compared to those obtained in the control group ($P = 0.25$).

Table 3 shows the comparative overall survival. It appears evident that the addition of prolonged combination chemotherapy to radical

Table 1

INFLUENCE OF LOCAL-REGIONAL VERSUS SYSTEMIC ADJUVANT THERAPY UPON RELAPSE AT 3 AND 4 YEARS

PRIMARY TREATMENT	YEARS ENTERED ON STUDY	No. OF PATIENTS	FIRST TREATMENT FAILURE (%)		
			3 yr.	4 yr.	
RM	1964-67	381	51.7	58.8	P= 0.06
RM + RT	1968-72	294	45.7	51.0	
RM vs	1973-75	179	47.8	52.7	P< 0.0001
RM + CMF		207	30.4	34.4	

RM: radical mastectomy

RT: postoperative radiation therapy

Table 2

COMPARATIVE INFLUENCE OF MENOPAUSAL STATUS
UPON FIRST TREATMENT FAILURE (%) AT 4 YEARS

	TOTAL	LOCAL REGIONAL	DISTANT ± LOCAL-REGIONAL
PREMENOPAUSE			
RM	55.2	10.7	44.5
RM + RT	56.8	7.9	48.9
RM vs	59.2	10.8	48.4
RM + CMF	25.0	4.9	20.1
POSTMENOPAUSE			
RM	60.0	18.3	41.7
RM + RT	48.4	9.0	39.4
RM vs	47.6	9.5	38.1
RM + CMF	43.8	8.4	35.4

RM: pre vs post P= 0.32
RM+RT: pre vs post P= 0.08
RM (control) pre vs post P= 0.17
PREMENOPAUSE: RM (control) vs CMF P= 0.00001
POSTMENOPUASE: RM (control) vs CMF P= 0.25

Table 3

COMPARATIVE OVERALL SURVIVAL AT 4-YEARS
FROM RADICAL MASTECTOMY (Actuarial analysis)

PRIMARY TREATMENT	YEARS ENTERED ON STUDY	OVERALL SURVIVAL (%)
RM	1964-67	62.2
RM + RT	1968-72	69.4
RM vs	1973-75	73.6*
RM + CMF		83.0*

* P = 0.05

mastectomy has provided the highest 4-year survival rate. It should also be noted that the overall survival was gradually improved over the years.

An important factor affecting survival could be the type of treatment(s) applied at the time of first relapse. Once primary treatment had failed, patients were usually subjected in sequence to numerous forms of therapy. Treatments were related, in the large majority of cases, to disease presentation, menopausal status as well as to changes in therapeutic concepts and drugs available at that particular time. Therefore, from this point of view, it appears difficult to adequately compare on a retrospective basis four series treated over a decade during which progress in medical treatment has been so dramatic. In fact, from 1970 combination chemotherapy was progressively utilized in our Institute for patients with advanced breast cancer (2). Table 4 shows both the percentage of different forms of therapy applied at first relapse and the comparative survival of relapsed patients. From the data reported, it cannot be denied that a more systematic use of effective chemotherapy applied at the time of first treatment failure gradually improved the survival.

In conclusion, at the Milan Cancer Institute, adjuvant chemotherapy with CMF was found superior in improving the early and intermediate course of operable breast cancer with positive axillary nodes when compared to radical mastectomy alone, radical mastectomy followed by postoperative radiotherapy and radical mastectomy plus adjuvant castration (2).

Adjuvant treatment was fairly well tolerated and no increased incidence of CMF-induced second neoplasms was so far observed. However, while in premenopausal patients 12 cycles of CMF after conventional surgery is to be considered as the treatment of choice, in postmenopausal women adjuvant chemotherapy remains experimental, until more solid data become available.

(Supported in part by Contract NO1-CM-33714 with DCT, NCI, NIH).

References

1. Bonadonna G., Brusamolino E., Valagussa P., et al.: Combination chemotherapy as an adjuvant treatment in operable breast cancer.
N. Engl. J. Med. 294: 405-410, 1976.

2. Bonadonna, G., Valagussa, P., Rossi, A., et al.: Are surgical

Table 4

TYPE OF FIRST TREATMENT (%) AT FIRST RELAPSE AND
COMPARATIVE SURVIVAL OF RELAPSED PATIENTS (%)

PRIMARY TREATMENT	NO. RELAPSED WITHIN 4 YRS	CC	EM	RT	SURVIVAL YEARS AFTER FIRST RELAPSE 1	2	3
RM	224	-	55	27	52.8	30.1	16.7
RM + RT	151	11	54	13	58.4	39.7	23.7
RM vs	88	59	26	7	81.3	56.9	39.4
RM + CMF	64	52	31	3	65.8	51.9	31.4

RM (control) vs RM + CMF: P = 0.19

CC: Combination chemotherapy
EM: Endocrine manipulations
RT: Palliative radiation therapy

adjuvant trials altering the course of breast cancer? Semin.Oncol. 1978 (in press)

3. Fisher, B., Carbone, P., Economou S.G., et al. : L-phenylalanine mustard (L-PAM) in the management of primary breast cancer. A report of early findings. N. Engl. J.Med. 292: 117-122, 1975.

4. Fisher, B., Slack, N., Katrych, D.L., et al.: Ten year follow-up results of patients with carcinoma of the breast in a cooperative clinical trial evaluating surgical adjuvant chemotherapy. Surg. Gynecol. Obstet. 140: 528-534, 1975

5. Fisher, B.: Biological and clinical considerations regarding the use of surgery and chemotherapy in the treatment of primary breast cancer. Cancer 40: 574 - 587, 1977.

6. Valagussa, P., Bonadonna, G., Veronesi U.: Pattern of relapse and survival following radical mastectomy. Analysis of 716 consecutive patients. Cancer 41: 1170-1178, 1978.

Comments to Dr. Valagussa's paper

1

Dr. Ribeiro

It was suggested to Dr. Valagussa that if she was now starting a trial of new adjuvant cytotoxic therapy for postmenopausal women then she should have a control group and not use historical controls.

2

Dr. Baum

I cannot understand the logic of this new Milan trial as regards post-menopausal women. Having failed to demonstrate an advantage for adjuvant chemotherapy in node-positive post-menopausal women, why are you now dropping a control group in the new phase of the study and comparing CMF versus a new more aggresive chemotherapy regime? Surely it would be more logical to compare a control with a new chemotherapy regime.

Answer to Dr. Ribeiro and to Dr. Baum

3

Dr. Valagussa

There is no need to have a new series of control patients treated only with radical mastectomy (RM) as we can derive the appropriate control group from our previous program comparing RM vs RM + 12 CMF cycles. The use of 'historical controls' is appropriate under certain circumstances. The characteristics of postmenopausal patients (eligibility criteria) have remained the same both in the new ongoing study and in the previous protocol with the exception of the age. Furthermore, from the experience achieved in our institute during the last 15 years, there is no change in the relapse-free survival of operable breast cancer patients with positive axillary nodes.

Problems in Withdrawal of Patients in a Randomized Study, when Treatment in one of the Groups cannot be carried out .

Sigvard Kaae and Helge Johansen
Aarhus, Copenhagen

In a randomized study carried out at the Radium Centre in Copenhagen from November 1951 to 1957, a comparison was made between simple mastectomy with postoperative X-ray irradiation (McWhirters method) and extended radical mastectomy (Dahl-Iversen's method)(1). A comparison of the two methods showed the same survival as well as recurrence-free survival up to 15 years for operable cases as a whole as well as for clinical stage I and for clinical stages II + III, although the number of patients in the latter group is too small to draw certain conclusions. The incidence of local/regional recurrences was higher after extended radical mastectomy without postoperative irradiation than after simple mastectomy plus postoperative irradiation in the locally and regionally more advanced, but still operable patients. The postoperative mortality was higher following extended radical mastectomy.

Among 206 clinical operable cases planned to have extended radical mastectomy the operation could not be carried out in 25 cases or 12%. In 15 cases the operation could not be carried out for technical reasons. There were metastases in the top of the axilla or in the supraclavicular region fixed to the vessels, and in one case, metastases to the internal mammary chain with diffuse infiltration in the surroundings. 8 patients were in too poor a condition for extended radical mastectomy, and two refused to have extended radical mastectomy. Instead, the patients had simple mastectomy, in some cases with partial excision of the lymph nodes and postoperative X-ray irradiation. These cases are included in the extended radical mastectomy group in the comparison with the results after simple mastectomy with postoperative X-ray irradiation, where the treatment in all 219 cases could be carried out.

In the more advanced clinical operable cases, clinical stages II + III, it was more frequent that extended radical mastectomy could not be carried out due to the infiltration of the metastases in the top of the axilla, in the supraclavicular region or in the internal mammary chain. Among 65 patients, 17 proved inoperable at the operation for this reason.

The table and figure show the crude survival rates up to 10 years. There are no differences between the results after the McWhirter method and after extended radical mastectomy including the inoperable cases (planned extended radical mastectomy). The results after extended radical mastectomy excluding the cases that proved inoperable at operation is much better than 10 year survival in 38% compared with 29% in the McWhirter group as well as in the 'planned extended radical mastectomy' group. Although this is a special and obvious situation, there are many similar in randomized studies.

Today a common comparison is between operation with and without supplementary chemotherapy. Here the operations can be carried out before randomization. There is no problem in the group without supplementary chemotherapy. In the chemotherapy group some patients refused chemotherapy at all, others refuse to continue chemotherapy under the course due to the side effects and in some cases the chemotherapy can not be carried through due to severe complications.

In some reports, a comparison is made between the group planned to have no chemotherapy with the patients, where chemotherapy has been carried out after the plan or at least to a described extent.

In doing so there is a risk of selection in the chemotherapy group. In particular the cases that do not tolerate chemotherapy well may be cases with a poor prognosis. The correct procedure must be to compare the group with no supplementary chemotherapy with the total group of patients planned to have supplementary chemotherapy. This will also give a comparison of the two treatment regimes as they can be carried out. Thereafter one may look at the cases where chemotherapy has been carried through compared with the other group, as well as the cases where chemotherapy could not be given as planned.

Table 1

Stages II + III

	No. of Cases	Crude Survival Rate	
		5 year	10 year
McWhirter's Method	70	46%	29%
Planned ext. Radical Mastectomy	65	48%	29%
Ext. Radical Mastectomy	48	54%	38%

STAGES II + III

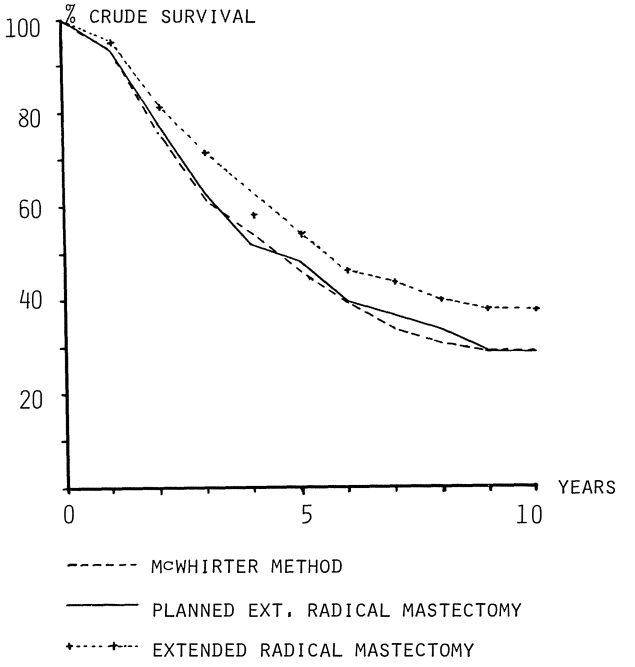


Figure 1 Crude survival rates in operable carcinoma of the breast, stages II + III, after simple mastectomy plus postoperative X-ray irradiation (McWhirter-method), after planned extended radical mastectomy (Dahl-Iversen's method) and after extended radical mastectomy excluding cases proved inoperable at operation.

Literature

Kaae, S. and H. Johansen: Does simple mastectomy followed by irradiation offer survival comparable to radical procedures? Int J. Radiation Oncology Biol.Phys. 1977, 2:1163-1166

Comments to Dr. Kaae's paper

1

Dr. Haybittle

I feel that the first question to be answered in a clinical trial is "Is there any difference in results obtained between two treatment policies?" If patients allocated to one arm do not complete treatment or refuse treatment, then this may well be because of some deficiencies inherent in the treatment policy of that arm, and the patients must be retained and included in the analysis. In comparing the policies, I do not think it matters how large a fraction of patients fall in this category.

2

Dr. Crowley

I agree that protocol deviants should be included in the analysis for answering the question of which treatment policy is best. On the other hand, there may be additional scientific questions which can best be addressed by excluding certain people who have not been properly treated. Of course, the possible bias should be considered carefully.

3

Dr. Baum

I merely wish to support Dr. Kaae's position that patients cannot be withdrawn from trials once treatment allocation is completed. Evev if the patient does not go on to receive the treatment allocated by random procedures, that patient must continue to be followed up as if within that treatment category. The fact that a number of patients will not receive the allocated treatment is as much a reflection of the strategy under investigation as the treatment itself.

With regard to informed consent, I think it is a naive hope that there

is anything that can be considered as truly informed consent. There can be very few patients, other than those who are themselves medically qualified, who can understand the implications of a prospective randomized clinical trial, nor can possibly understand the possible risk benefit of any treatment allocation. I therefore think that the British attitude concerning 'informed consent' is not only ethical but compassionate in dealing with patients with potentially fatal disease. Seeking a spurious informed consent merely to protect the doctor from litigation is not in the patient's best interest.

Treatment departure and survival analysis in a randomized trial on the value of pre- and postoperative radiotherapy in breast cancer

Arne Wallgren, Britta Mattsson and Leif Karnström.
The Oncologic Centre, Radiumhemmet
Stockholm

Some trials on the value of various modalities of local or systemic treatment in breast cancer have been complicated by a large number of treatment deviations. When reporting the results of such trials the patients who had not been treated according to the protocol sometimes were excluded from the analyses, assuming that the reasons for the deviations from treatment were not of a nature which could bias the results.

This paper will illustrate the problem, with the treatment deviations in a trial on the value of preoperative radiotherapy in comparison with surgery only and surgery plus postoperative radiotherapy in operable breast cancer.

The clinical trial

The study was undertaken as a cooperative trial in Stockholm 1971-1976. Participating departments were Radiumhemmet and the surgical departments of the Karolinska sjykhuset, Serafimerlasarettet, Sabbatsbergs sjukhus, S:t Eriks sjukhus and S:t Görans sjukhus. The treatment protocol and preliminary results are presented elsewhere (1,2) and will only be summarized here.

Female patients less than 71 years of age with unilateral , operable breast cancer were considered eligible for the study if the diagnosis of breast cancer was confirmed preoperatively by means of fine needle aspiration biopsy. The total number of patients in the study was 960. Of these 316 had been randomized to receive radiotherapy preceding a modified radical mastectomy. The remaining 644 patients constituted the control group in which the treatment consisted in the same surgical procedure. 323 of these patients were randomly allocated to get post-operative radiotherapy and 321 patients to receive no further treatment.

The radiotherapy was individually planned, whether given preoperatively or postoperatively, to give a dose of 4,500 rad in about 5 weeks to

the breast and the chestwall, the internal mammary, the supraclavicular and the axillary lymph nodes.

None of the patients have been lost to follow-up. All patients have a trial time of at least one year and 356 patients of at least five years.

Both preoperative and postoperative radiotherapy significantly increase the recurrence-free survival rates compared to those treated by surgery only (log-rank test, p=0,0001), mainly through the reduction of the incidence of local or regional recurrence of the disease. There was no difference in this respect between preoperative and postoperative radiotherapy.

The preoperatively irradiated patients had a significantly increased survival rate compared to the two control groups (p=0,05); if analysed separately, preoperative irradiation significantly increased the survival time compared to only surgically treated patients (p=0,03) but not compared to postoperatively irradiated patients (p=0,20). No obvious difference was found between the two control groups (p=0,35).

Analyses have been performed of the therapeutic effect in various clinical subsets of patients, i.e. according to menopausal status, the clinical size of tumour, the clinical assessment of lymph nodes and the site of tumour in the breast. The preoperatively irradiated patients showed lower death rates in all subsets compared to the two control regimes. Only in patients with tumours located to the inner half of the breast, with tumours less than 3 cm, or with clinical uninvolved axillary lymph nodes, did preoperative irradiation give significantly lower mortality ($p = < 0,05$) than radical mastectomy only. However, analyses of many subsets of patients are likely to produce spurious differences, so variations in therapeutic effect between such subsets should be judged cautiously.

Treatment departures

Major deviations from the treatment as defined in the protocol occurred in only 34 (3,5 per cent) of 960 patients in this trial. Table 1 gives the causes for protocol violations.

The main cause for deviations from treatment among the patients allocated to preoperative radiotherapy was the discovery of distand metastases during the delay of surgery for more than three months caused by the preoperative irradiation. Consequently, these patients were not surgically treated. Three patients refused mastectomy after radiothe-

Table 1. Deviations from treatment

Treatment groups Causes for treatment	No. of patients
1. Patients randomized to preoperative radiotherapy	
No treatment; patient refuses	1
No radiotherapy; pulmonary insufficiency	1
No surgery; distant metastases	5
No mastectomy; patient refuses	3
Total No. of deviants	10
2. Patients randomized to postoperative radiotherapy	
No mastectomy; patient refuses	1
No radiotherapy; patient refuses	8
No radiotherapy; distant metastases	3
Total No. of deviants	12
3. Patients randomized to surgery only	
No treatment; patient refuses	2
No radical mastectomy	2
Radiotherapy given; non-radical surgery	8
Total No. of deviants	12

rapy, presumably as a result of the shrinkage of the tumour.

Among the patients randomized to receive radiotherapy after the modified radical mastectomy, distant metastases became evident in three cases before the initiation of radiotherapy, and these patients were given systemic treatment instead of radiotherapy. In further two patients, not included among the deviants of Table 1, the postoperative radiotherapy was interrupted at a lower dose than prescribed in the protocol because of such metastases.

The most common cause for protocol violation in the group allotted to postoperative radiotherapy was, however, that the patients refused radiotherapy. None of these patients, who pursuaded the doctor not to give radiotherapy had involved axillary lymph nodes.

Eight of the patients who were to be treated by surgery only were postoperatively irradiated. Seven of these had involved nodes and surgery was not considered to be radical.

In all analyses the deviants have been retained in their original groups. Those who were not surgically treated after radiotherapy or who refused treatment were considered as local failures. Because of their small number, the withdrawal of the deviants from analyses would have yielded similar results as those reported. The probability that the difference in survival rates between the preoperatively irradiated patients and the control groups was caused by chance would have decrease from p=0,05 to p=0,02.

Conclusion

It is obvious from the reasons for protocol violations listed in Table 1 that few if any of the deviations have occurred at random and independent of the proposed treatment. If more numerous, exclusion of deviants is likely to bias the result severely.

References

1. de Schryver, A.: The Stockholm breast cancer trial: Preliminary report of a randomized study concerning the value of pre-operative or post-operative radiotherapy in operable disease.
Int.J.Radiat.Oncol.Biol.Phys. 1: 601-609, 1976.

2. Wallgren, A., Arner, O., Bergström, J.,Blomstedt, B., Granberg,P-O., Karnström, L., Räf, L., and Silfverswärd, C.: Preoperative radiotherapy in operable breast cancer. Results in the Stockholm breast cancer trial.
Cancer 42: 1120 - 1125, 1978

Comments to Dr. Wallgren's paper

1

Dr. Haybittle

Dr. Wallgren's result in medial tumours is surprising considering the number of other trials which have found no advantage for radiotherapy as far as survival is concerned. It is very difficult to give adequate radiotherapy dosage to the internal mammary nodes. Could it be that Dr. Wallgren's technique was more successful at doing this than were the techniques used in the other trials?

2

Dr. Roberts

I would like to question Dr. Wallgren on the contribution of node pathology to the interpretation of survival data. I believe you said that the number of patients with node histology in the pre-operatively irradiated group was low, but those with negative nodes had no difference in survival between the 3 treatment groups, whereas if the nodes were positive, then patients treated with pre-operative irradiation survived better than the other two groups.

As a consequence I think the pre-operative irradiation policy should not be given in patients with lateral tumours, in view of the fact that much valuable information such as node histology and oestrogen receptor data would be lost, and patients derived no benefit from this form of treatment.

3

Dr. Stewart

I feel that reduction in node positive cases in the pre-operatively irradiated group is a most interesting finding in the results just presented. Could it not represent a pointer in explaining the advantage suggested in the overall survival results for the pre-operatively irradiated medial tumour group? Timing of treatment as well as site of treatment are different in all three groups. I think it might be revealing to look at the overall distribution of axillary node status in terms of site of tumour (medial and lateral) for all three treatment options despite the smallness of the resulting groups.

4

Dr. Ribeiro

It is difficult to explain why the group with medial tumours did worse but there is a possibility that the internal mammary nodes were not adequately irradiated. It would be important to know if the axillary nodes were also histologically positive in these patients but this will never be possible as the axillary has already been irradiated pre-operatively.

5

Dr. Baum

I have the greatest difficulty of fitting your findings into any biological model that could describe the natural history of carcinoma of the breast. However, to get any difference in survival according to different protocols of local therapy is in itself of great interest as there is a growing tendency throughout the world for clinicians to make the assumption that the extent of local therapy no longer influences the chance of survival.

Stochastic Models in Clinical Trials*

Stephen W. Lagakos, Ph.D.

Department of Biostatistics
Harvard School of Public Health
and

Sidney Farber Cancer Institute
Boston

1. Introduction

In most analyses of clinical trials, the therapies being investigated are assessed on the basis of several time-dependent events. For example, in cancer clinical trials events such as drug toxicity, disease relapse, tumor remission, and death are common measures of therapeutic effect. These events are time-dependent in the sense that each can occur at various points in time after initiation of treatment. For those such as death which are certain to eventually occur, the interest is in the time until the event. For events such as tumor remission, which may or may not occur, both frequency and time until the event are of interest. Furthermore, knowledge of the relationships between events is often valuable. For example: How is remission related to survival? Does an elevated tumor marker signal impending failure? Is metastatic disease associated with early death?

A common practice in the evaluation of clinical trial data is to analyze each of several endpoints separately. Although this approach is adequate in some situations, there can be other instances where alternative approaches are preferable. First, there can be useful information obtained from an assessment of the correlation between two or more endpoints. Secondly, it is often possible to obtain more informative estimates of the parameters of a specific endpoint by taking account of its relationship with other endpoints. Finally, the simultaneous rather than separate analysis of several endpoints usually provides a better 'feel' for an individual patient's entire experience, and the knowledge that outcomes correspond to some probabilistic law can sometimes be useful in designing studies.

---

*This investigation was supported by Grant Number CA-00505 and CA-23415 awarded by the National Cancer Institute, DHEW.

An alternative approach to the separate analysis of individual events of interest is to model the patient experience as a multi-state stochastic process. To illustrate, consider a clinical trial where the events of interest are disease remission, objective (measurable) disease progression, and death, and where the set of possible patient paths are depicted in Figure 1.

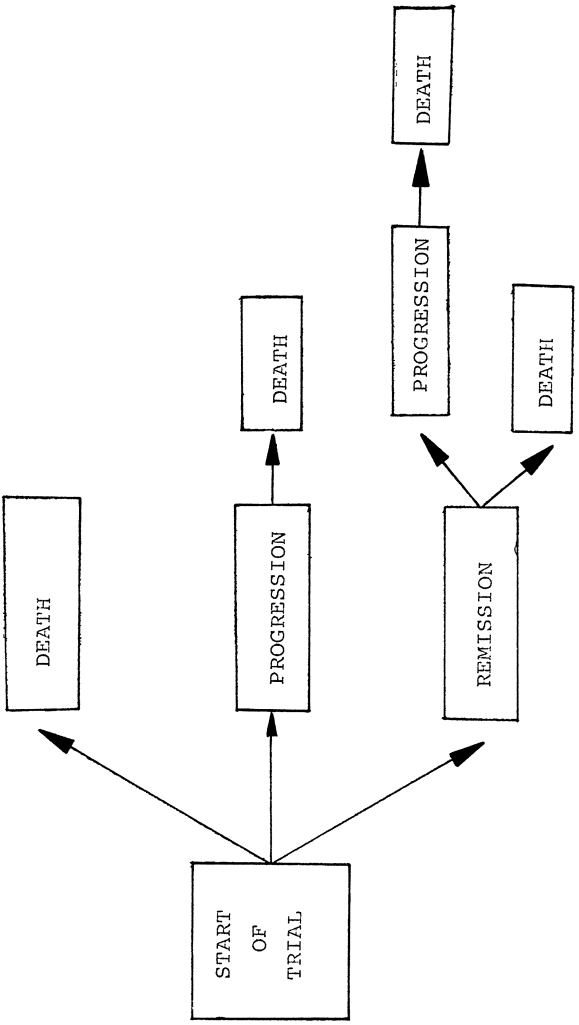
Each patient starts in a state of active disease and then experiences one of three events: remission, objective progression, or death (without remission or progression). Those patients who experience a remission must next either experience a progression or a death without objective progression. Those patients who initially experience a progression can next only expire.

A 'standard' analysis of data of this type might consist of analysis of remission rates, time to remission, time to progression, and time to death. Alternatively, a stochastic-process approach would represent each event as a state and, from this, individual analyses would arise as specific components of the process. For example, in the phraseology of stochastic processes, survival time corresponds to the first passage time to the state of death and tumor remission incidence corresponds to a transition probability from the initial state of active disease.

Although the formulation of clinical-trial data in terms of stochastic processes seems natural and conceptually appealing, the use of stochastic models in clinical trials has been extremely limited. The primary reason for this is complexity, resulting from the fact that any such model must describe the probabilistic properties of the entire patient experience. Clearly, this will usually require a more complicated model than one based only on a single endpoint such as survival time. A second complicating factor is the presence of right censored observations. These arise when some patients are still under study at the time of analysis, and therefore yield only 'partial histories'. Although there are a great number of existing stochastic models for possible use in clinical trials, very few have been formulated to accomodate censored data.

The purpose of this paper is to discuss the use of one family of stochastic processes in clinical trials--namely, semi-Markov or Markov renewal models. The intrinsic semi-Markov model is quite simple, and has been found to accurately represent several disease sites in cancer clinical trials and probably has wide application elsewhere. In addition, the incorporation of censored observations

Figure 1.


START OF TRIAL
DEATH
PROGRESSION
DEATH
REMISSION
PROGRESSION
DEATH
DEATH

presents no great difficulty and so the use of semi-Markov models in analyses is quite feasible, although several other technical questions still remain to be answered. Our purpose is not so much to advocate a specific statistical method, but instead to present a set of data and a different approach to its investigation.

The remainder of this paper is divided into three sections. In Section 2, we give our notation and introduce the usual semi-Markov model. Section 3 illustrates the fitting of the model to a set of data from a clinical trial for small cell carcinoma of the lung. Section 4 discusses some technical considerations which require further investigation. For a more thorough discussion of the topics considered in this paper, see Lagakos, Sommer, and Zelen (1978) and Lagakos and Zelen (1978). For examples of other stochastic models for partially censored data, see Hanley and Parnes (1978), Turnbull, Brown and Hu (1974), and Crowley and Hu (1977). Also related is the use of time-dependent covariates in models for survival data [cf: Prentice, et al. (1978)].

2. Semi-Markov models

Let us first consider the experience of an individual patient. Suppose that at each point in time the patient is in one of s states denoted 1,2,...,s. The entrance into a state might correspond to the occurrence of an event such as remission, death, etc,. and we can regard a patient as being 'in' this state until his or her next critical event occurs. Without loss of generality the first s_1 states are assumed to be transient (i.e., states to which return visits are not certain) and the last $s-s_1$ are absorbing (i.e., states from which there can be no escape). In the example, remission represents a transient state and death an absorbing state. A subject's history, therefore, consists of a sequence of events or epochs in time and ends once an absorbing state in reached.

Suppose Z_0 denotes the patient's initial state and Z_n is the state corresponding to the n^{th} epoch. Furthermore, let T_n represent the 'sojourn time' between the $(n-1)^{st}$ and n^{th} epoch. Then the entire patient history can be represented by

$$H = \{Z_0, T_1, Z_1, T_2, Z_2, \ldots, T_m, Z_m\} \quad , \tag{1}$$

where, by definition, $1 \leq Z_i \leq s_1$

for $i < m$ and $Z_m > s_1$. To illustrate, consider Figure 1. and suppose the initial state, remission, progression, and death states are

labeled 1,2,3, and 4. Then $s_1 = 3$ and $s = 4$. The history of a patient who first experiences a remission and next dies is thus of the form

$$\{1,t_1,2,t_2,4\}$$

where t_1 is the time until remission and t_2 is the time from remission to death. Note that the overall time to death is $t_1 + t_2$.

In order for the process governing H to be a semi-Markov process [cf:Cox and Miller (1965)], two conditions must be satisfied. First, the sequence of epochs $\{Z_0,Z_1,\ldots,Z_m\}$ must form a homogeneous Markov chain. This means that the patient's next state depends only on his current state and not on any previous states. Secondly, the sojourn times between epochs must be independent and depend only on the adjoining states. This means, among other things, that the duration of time in a given state does not depend on either the time needed to reach that state or the previous state.

When the process is semi-Markov, its probabilistic properties are determined by the quantities $\theta(i)$, $\theta(i,j)$, and $Q(t;i,j)$, where

$$\theta(i) = P[Z_0 = i] \tag{2}$$

is the probability that the patient is initially in (transient) state i,

$$\theta(i,j) = P[Z_{n+1} = j | Z_n = i] \tag{3}$$

is the conditional probability that the next state is j, given that the current state is i, and

$$Q(t;i,j) = P[T_n > t | Z_{n-1} = i, Z_n = j] \tag{4}$$

is the conditional probability that the sojourn time between the $(n-1)^{st}$ and n^{th} epoch exceeds t, given that these epochs correspond to states i and j, respectively. It can be easily shown that the probability associated with the history H in (1) is given by

$$\theta(Z_0) \prod_{n=1}^{m} \{\theta(Z_{n-1},Z_n)\ Q'(t_n;Z_{n-1},Z_n)\}\ , \tag{5}$$

where Q' is the derivitive of Q.

Returning to Figure 1, we see that

$$\theta(1) = 1,\ \theta(i) = 0 \quad \text{for} \quad i \neq 1, \quad \theta(i,j) = 0 \quad \text{for} \quad i \geq j,$$

and $\theta(3,4) = 1$. Also, the four possible patient paths and their association probability elements are given in Table 1.

Table 1. Paths and Probabilities for Example Depicted in Figure 1.

Path	H	Probability Element
Death at time t_1	$\{1,t_1,4\}$	$\Theta(1,4)Q'(t_1;1,4)$
Progression at time t_1, death at time t_1+t_2	$\{1,t_1,3,t_2,4\}$	$\{\Theta(1,3)Q'(t_1;1,3)\}\{Q'(t_2;3,4)\}$
Remission at time t_1, death at time t_1+t_2	$\{1,t_1,2,t_2,4\}$	$\{\Theta(1,2)Q'(t_1;1,2)\}\{\Theta(2,4)Q'(t_2;2,4)\}$
Remission at time t_1, progression at time t_1+t_2, death at time $t_1+t_2+t_3$.	$\{1,t_1,2,t_2,3,t_3,4\}$	$\{\Theta(1,2)Q'(t_1;1,2)\}\{\Theta(2,3)Q'(t_2;2,3)\}\{Q'(t_3;3,4)\}$

In order to apply these models to a set of data, one must estimate the unknown quantities $\theta(i)$, $\theta(i,j)$, and $Q(t;i,j)$. Sometimes it may be desirable to further model each $Q(t;i,j)$ in terms of a parametric family of distributions. For example, one could take $Q(t;i,j) = \exp\{-\lambda_{ij}t\}$. which means that the sojourn time between states i and j is exponentially distributed with rate parameter λ_{ij}. A second consideration is the method used for estimating $\theta(i,j)$ and $Q(t;i,j)$. Since some of the data is censored, many of the commonly used estimation methods do not directly apply. One that does, however, is the method of maximum likelihood, and this is the technique that is used for the estimates presented in the next section.

3. An Example

In this section the use of semi-Markov models is illustrated with data from a clinical trial for inoperable small cell carcinoma of the lung conducted by the Eastern Cooperative Oncology Group [see Lagakos and Zelen (1978) for further consideration of these data]. All patients entered the study in a state of active disease and experienced a remission, progression, or death as depicted in Figure 1.

Table 2 gives the histories of the 70 patients in the trial who received one of the treatments being investigated. All sojourn times have been rounded-off to the nearest week and state 5 (= s + 1) is used to denote a censored observation. Thus, for example, patient 1 died in 21 weeks without a progression or remission, while patient 22 experienced a progression after 8 weeks and is still alive 11 weeks later.

Of the 70 patients, 27 initially experienced a remission, 21 initially progressed, 20 died without a progression or remission, and 2 were censored before any event occurred. Nineteen of the 27 remissions were followed by progressions, 4 were followed directly by death, and 4 were censored. Of the 40 observed progressions (21 of which were initial events and 19 of which followed remissions), death was subsequently observed in 27 cases and the remaining 13 were censored.

Using the methods in Lagakos, Sommer, and Zelen (1978), the estimated transition probabilities from state 1 and associated standard errors are $\hat{\theta}(1,2) = .386 \pm .059$, $\hat{\theta}(1,3) = .300 \pm .055$, and $\hat{\theta}(1,4) = .314 \pm .056$. Thus, the estimated remission rate for this population of patients is 38.6%. The estimates of the corresponding sojourn-time distributions are depicted in Figure 2. Note that among initial events, remissions and progressions tend to occor sooner than deaths without remission or

Table 2 - Observed Histories (in weeks) of 70 Patients with Small Cell Carcinoma of the Lung

Patient	T_1	z_1	T_2	z_2	T_3	z_3	Patient	T_1	z_1	T_2	z_2	T_3	z_3
1	21	4					36	6	3	0	5		
2	12	4					37	3	3	1	4		
3	3	2	8	4			38	2	4				
4	6	3	0	5			39	6	3	30	4		
5	3	2	3	3	7	4	40	4	3	10	4		
6	7	4					41	4	2	41	5		
7	9	2	3	3	6	4	42	3	2	42	3	4	5
8	6	2	12	3	11	4	43	48	4				
9	32	5					44	16	4				
10	7	2	19	3	10	4	45	7	2	12	3	28	4
11	18	4					46	5	4				
12	10	4					47	12	2	9	4		
13	43	5					48	7	3	2	4		
14	9	3	1	4			49	10	2	11	4		
15	31	4					50	9	3	4	4		
16	11	3	36	4			51	14	4				
17	5	2	1	4			52	3	2	13	5		
18	2	2	13	3	6	4	53	9	3	15	4		
19	2	4					54	14	3	8	4		
20	3	2	17	3	19	4	55	10	3	0	5		
21	9	3	20	4			56	3	2	6	3	18	4
22	8	3	11	5			57	11	3	13	4		
23	20	2	12	3	1	5	58	3	2	2	3	1	5
24	12	2	16	5			59	6	2	6	3	10	5
25	1	4					60	3	2	26	5		
26	5	2	30	3	4	4	61	19	3	10	5		
27	9	2	31	3	8	5	62	7	3	7	4		
28	10	2	21	3	20	5	63	14	4				
29	3	2	9	3	2	5	64	6	4				
30	5	4					65	6	3	16	4		
31	13	2	14	3	12	4	66	13	3	8	4		
32	10	2	4	3	1	4	67	7	4				
33	20	4					68	9	2	3	3	0	4
34	7	3	3	4			69	6	4				
35	2	4					70	6	3	0	5		

Figure 2. Departures from Initial State

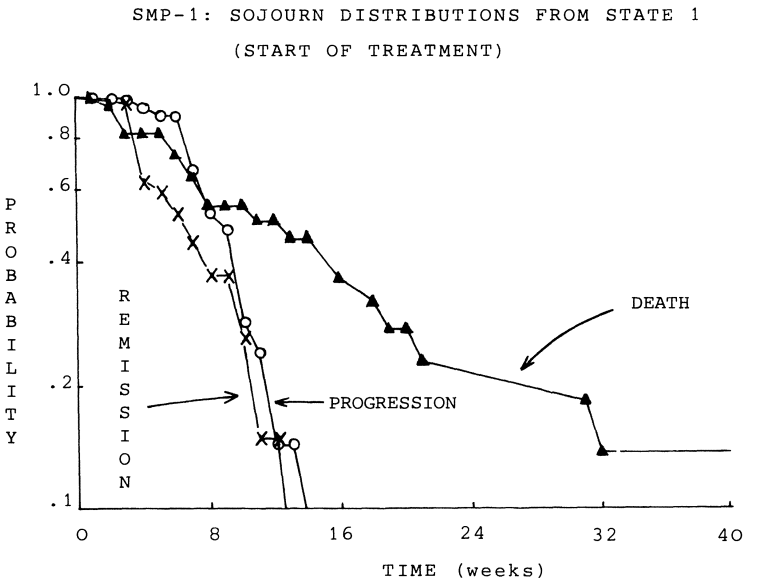


Figure 3. Events Following Remission

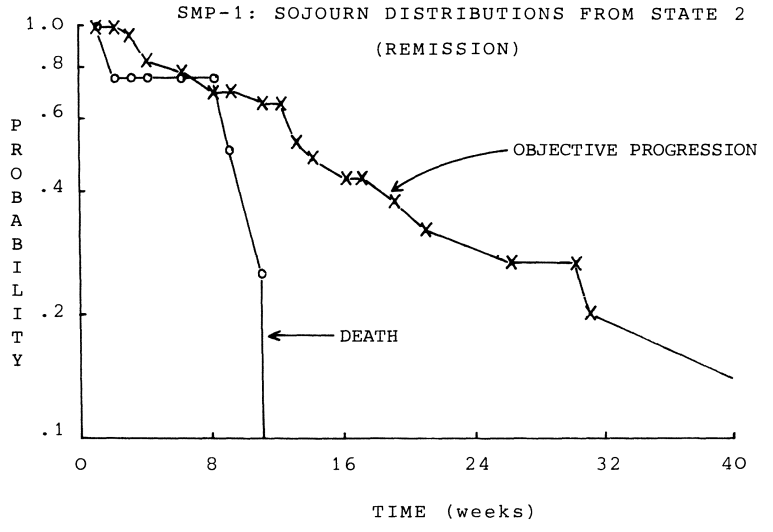


Figure 4. Time from Objective Progression to Death

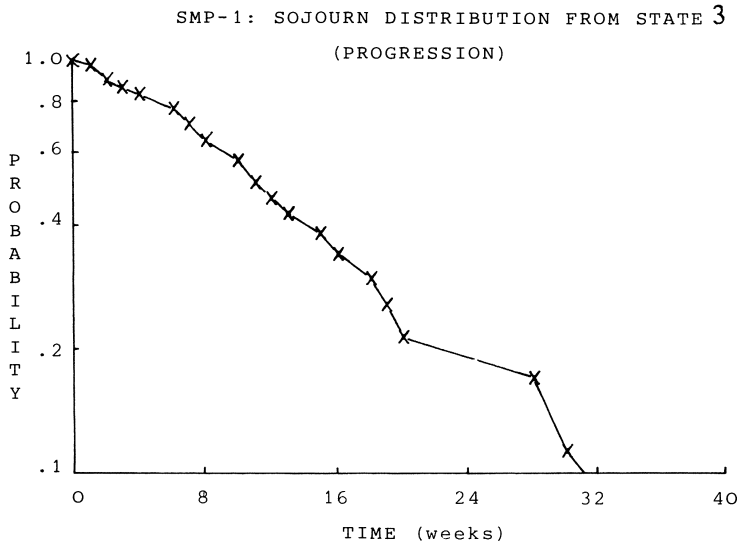


Figure 5 Patient Paths Indicating Transition Probabilities (in parentheses) and Median Sojourn Times

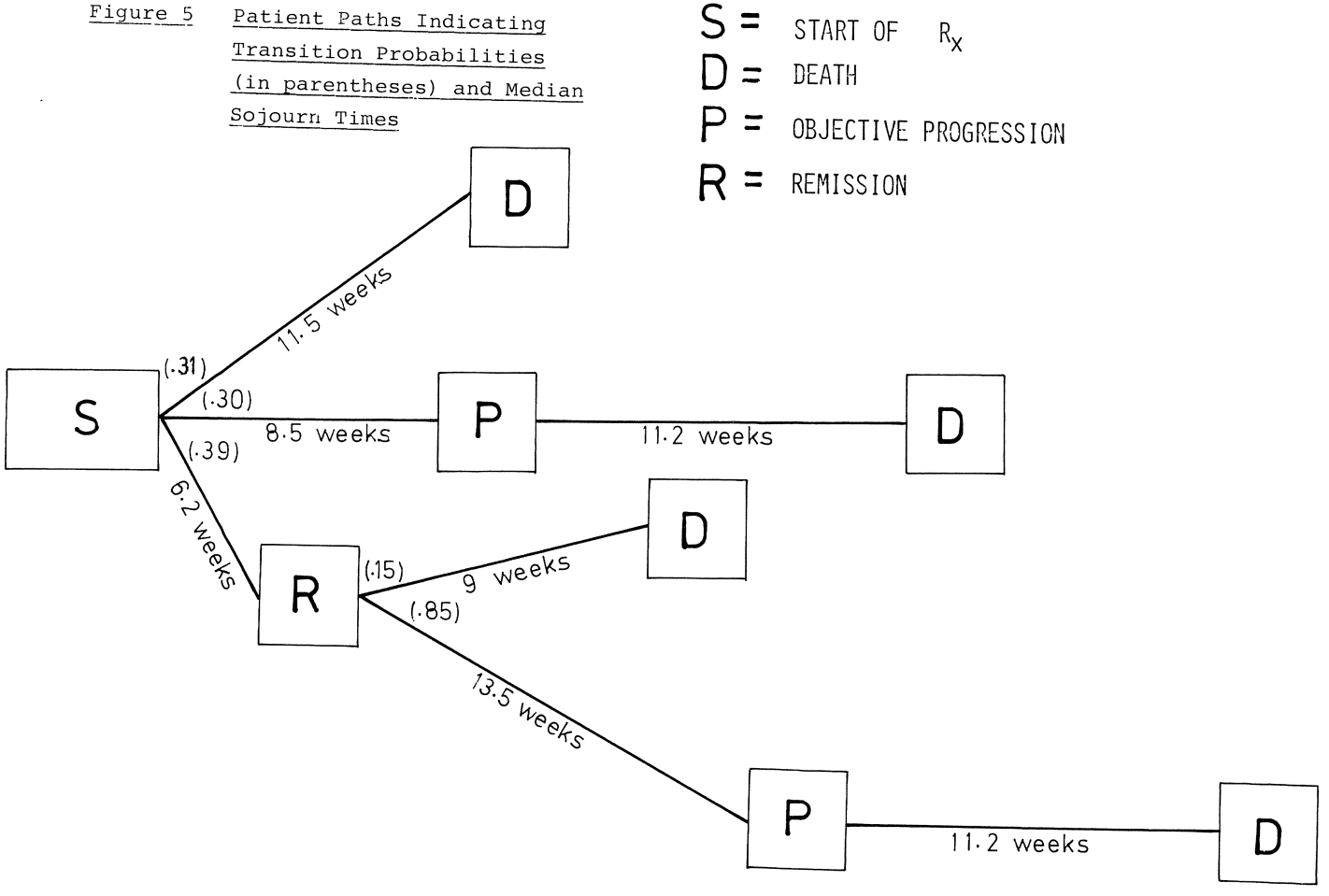
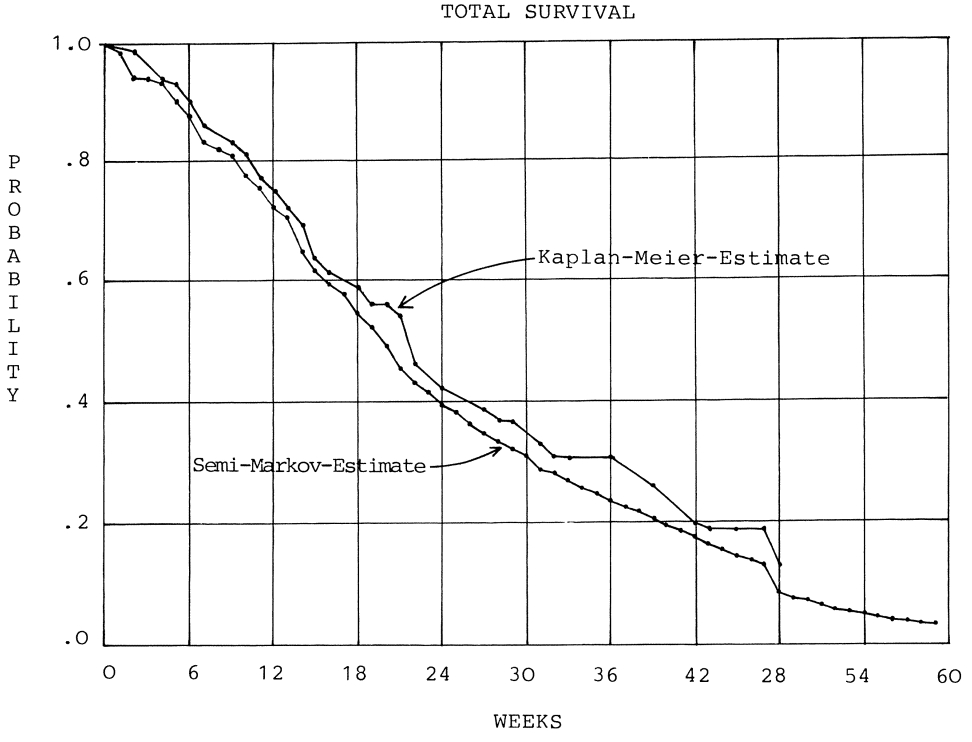


Figure 6. Kaplan-Meier and Semi-Markov Estimates of the Survivorship Function for Overall Time to Death



objective progression.

For patients who initially experience a remission, the estimated transition probabilities to progression and death are .852 ± .069 and .148 ± .069, respectively. The corresponding sojourn distributions are presented in Figure 3. Thus, death is preceded by objective progression, and when it is not, it tends to occur fairly soon in a great majority of cases. Finally, Figure 4 gives the estimated sojourn time from progression to death.

The preceding results are also summarized in Figure 5, where distances between successive states are proportional to the corresponding medians and transition probabilities are noted. This figure clearly indicates (1) that initial remission and progression, should they occur, tend to occur quickly, and (2) that most remissions are followed by progressions, but when they are not, that death occurs quite soon.

The usual methods for estimating the distribution of time to death are based only on the overall times to death and not on the times to intermediate events such as remission or progression. An alternative approach can be obtained by noting that overall survival is simply the first passage time from state 1 to state 4. Accordingly, the survivorship function for survival time can be written as

$$
\begin{aligned}
dF(t) = {} & \theta(1,2)\,\theta(2,4) \int_0^t dQ(s;1,2)\, dQ(t-s;2,4) \\
& + \theta(1,2)\,\theta(2,3) \int_0^t \left[\int_0^s dQ(r;1,2)\, dQ(s-r;2,3) \right] dQ(t-s;3,4) \\
& + \theta(1,3) \int_0^t dQ(s;1,3)\, dQ(s;3,4) \\
& + \theta(1,4)\, dQ(t;1,4) \; .
\end{aligned}
$$

We can thus estimate $F(t)$ by seperately estimating each of the terms in the right-hand-side of this equation. Figure 6 depicts this estimate, denoted $\hat{F}$, as well as the Kaplan-Meier (1958) estimator of F based only on the overall survival times of each patient (denoted $\hat{F}_{km}$). The increased precision of $\hat{F}$ may be considerable, although this point requires further investigation.

We conclude with a few brief remarks on model testing. When nonparametric methods are used to estimate the unknown parameters, the only assumptions being made are that state changes form a Markov chain and that sojourn times are independent and depend only on the adjourning states. Informal checks on both of these assumptions can easily be

obtained by appropriate partitioning of the data and reapplication of the estimation methods. For example, the sojourn times from state 3 (progression) can be grouped according to the state preceding progression (remission or initial state). For each group, the sojourn time distribution to death can be estimated and, if the semi-Markov assumption is accurate, the two distributions should be comparable. Similar methods can be used to determine whether time until entrance into a state (e.g. remission) affects the duration of time in this state and/or the subsequent state. These techniques were applied to the lung cancer data and supported the semi-Markov assumption quite well.

4. Technical Considerations

The preceding section indicates some preliminary steps in an analysis of clinical trial data using a semi-Markov model. A complete assessment of this or similar data, however, would include additional analyses to those presented here. We now briefly discuss some of these where further statistical research is needed.

We remarked earlier that the estimate of total survival resulting from the semi-Markov model is likely to be more precise than one based only on each patient's time to death. Intuitively, this is clear because the former is based on the relationships between events and thus utilizes the occurrence of and times to these events. It is not clear, however, how much information is gained by the semi-Markov approach, and thus it dous not follow that the additional computational effort and assumptions required of this approach make its use worthwhile. For these reasons, a theoretical investigation of the increased precision as well as robustness of the semi-Markov estimates would be useful.

A somewhat related point is that the distribution theory associated with estimates of first-passage-time distributions is largely unexplored. However, the recent work of Aalen (1976) and Gill (1978) involving counting processes is closely related and holds great promise in this regard.

Another problem of considerable importance is the assessment of whether one state 'influences' or is 'related' to another. For example, how do we determine if remission is related to survival time, or whether an increase in CEA is a precursor of relapse of disease? These are questions for which both definitions of effect as well as cor-

responding statistical tests are needed.

Finally, methods appropriate for heterogeneous patient populations need to be explored. The present methods can accomodate situations where there are a few number of different patient types or strata. This is done by simply allowing these strata to correspond to different initial states. However, this becomes impractical for more than a few strata and more sophisticated covariate methods are needed. The recent work of Prentice, et al. (1978) suggests one approach to this problem.

Acknowledgements

I am grateful to James A. Hanley and Marvin Zelen for their comments, to Karen Uliss for typing the manuscript, and to the Eastern Cooperative Group for allowing me to use their data.

This work was supported by grants CA-00505 and CA-23415 from the National Cancer Institute.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes.
Annals of Statistics, 6, 701-26.

Crowley, J.J. and Hu, M. (1977).
Covariance analysis of heart transplant survival data.
Journal of the American Statistical Association, 72, 27-36.

Gill, R.D. (1978). Nonparametric estimation based on censored observations of a Markov renewal process.
Technical report SW 64/78, Stichting Mathematisch Centrum, Amsterdam, Netherlands.

Hanley, J.A. and Parnes, M. (1978)
Estimation of multivariate distribution in the presence of censoring. Submitted to a technical journal

Kaplan, E.L. and Meier, P. (1958)
Nonparametric estimation from incomplete observations.
Journal of the American Statistical Association, 53, 457-81.

Lagakos, S.W., Sommer, C.J., and Zelen, M. (1978): Semi-Markov models for partially censored data. Biometrika, 65, 2, 311-17.

Lagakos, S.W. and Zelen, M. (1978). The use of semi-Markov models in clinical trials. Submitted to a technical mournal.

Prentice, R., et al. (1978). The analysis of failure times in the presence of competing risks. Biometrics, 34, 541-54.

Turnbull, B.W., Brown, B.W. and Hu, M. (1974). Survivorship analysis of heart transplant data. Journal of the American Statistical Association, 69, 74-80.

Editor's note

In his paper as presented at the Symposium Dr. Lagakos made some general comments concerning the need to standardise the measurements of prognostic factors between different institutions involved in a trial. He gave as an example measurements of Oestrogen Receptor Status obtained from several centres. This led to the content of the following comments.

Comments to Dr. Lagakos' paper

1

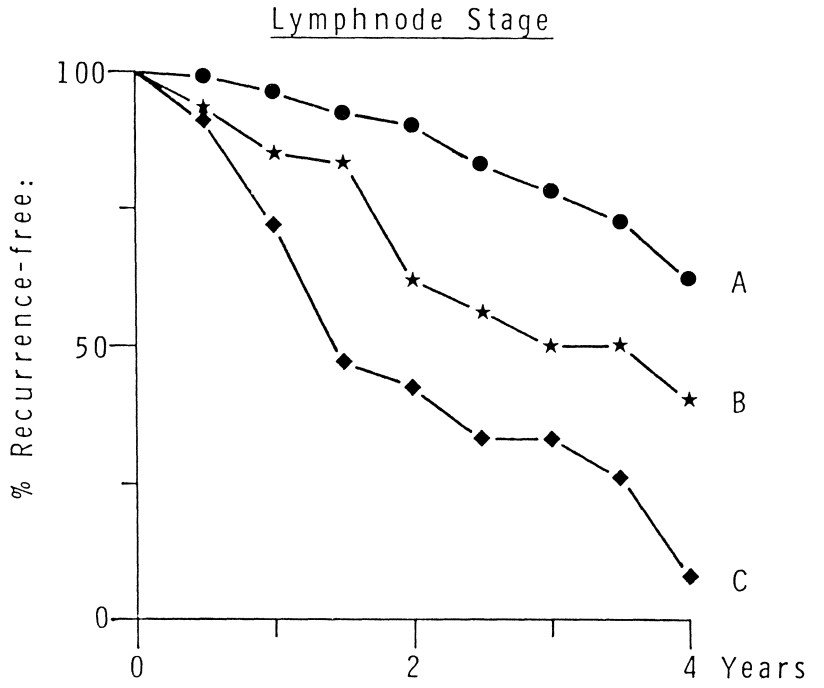
Dr. Haybittle

Dr. Lagakos said that the use of Oestrogen Receptor Status as a prognostic factor was still in some doubt. Some results from M.R. Blamey's group at the Department of Surgery, Nottingham City Hospital*, may be of some interest in this connection. They are derived from 196 consecutive patients seen in Mr. Blamey's Breast Clinic who were operated on by a simple mastectomy and, at operation, had biopsies taken of a pectoral node, an apical node adjacent to the axillary vein and a second intercostal space node lying within the internal mammary vessels. Patients were staged according to their lymphnode status as follows:

A - no lymphnode invasion
B - invasion of pectoral node
C - invasion of apical and/or internal mammary node.

Fig. 1 shows the considerable influence of lymphnode stage on prognosis measured as time to first recurrence. Oestrogen receptor assays were made by Professor K. Griffiths' group at the Tenovus Institute for Cancer Research at Cardiff, and fig. 2 shows that in the whole

Fig. 1 (Dr. Haybittle's Comment)



100	95	77	32	8	A
60	51	33	12	3	B
36	26	15	10	0	C

Log rank $X^2 = 38.02$; d.f. = 2; $P<0.0005$

Fig. 2 (Dr. Haybittle's Comments)

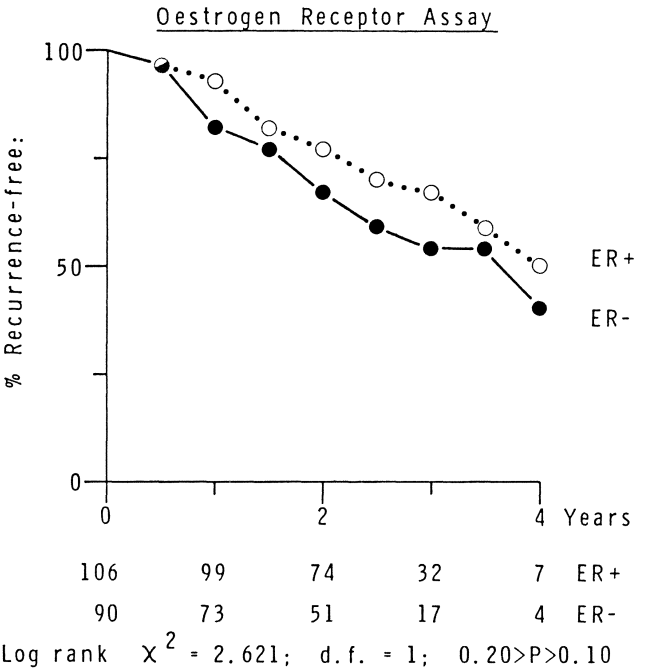
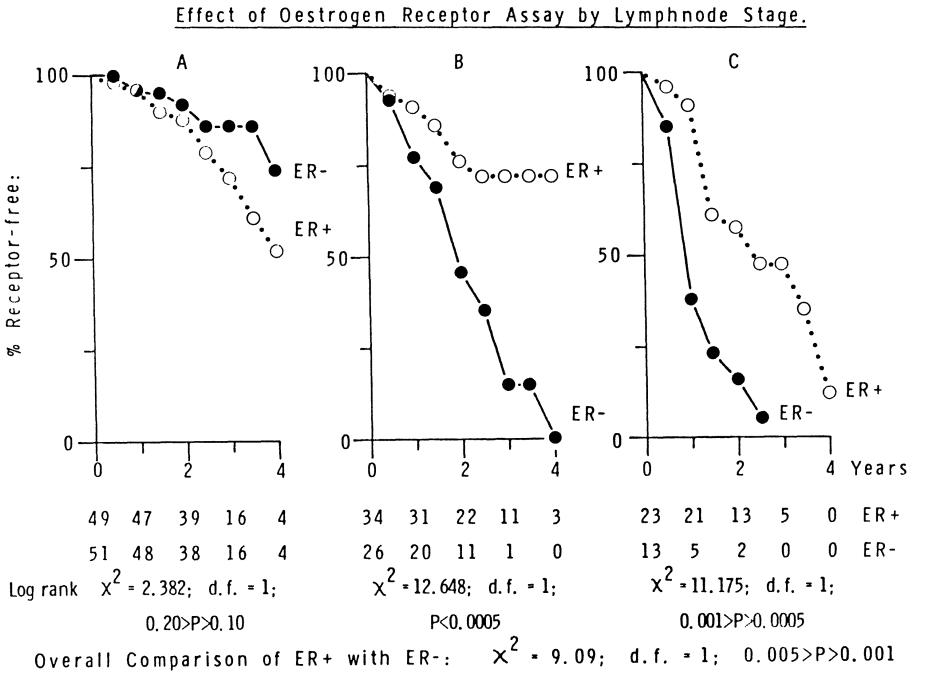


Fig. 3 (Dr. Haybittle's Comments)



---

*Maynard,P.V., Blamey, R.W., Elston, C.W., Haybittle, J.L., and Griffiths, K., (1978) Estrogen Receptor Assay in Primary Breast Cancer and Early Recurrence of the Disease. Cancer Research, 38, 4292-4295

group the results from ER+ and ER- patients were not very different although there is a tendency for ER+ patients to do better.

However, when ER+ and ER- patients are compared in the different lymphnode stage categories (fig. 3), it is apparent that ER- patients do markedly worse in lymphnode stage B and C, and the overall comparison is now highly significant. Oestrogen Receptor Status may therefore be an important prognostic sign in cases with lymphnode invasion.

2

Dr. Roberts

You may be interested in the results we published recently in the British Journal of Cancer (September, 1978) describing the interchange of tissue and cytosol samples between 5 different laboratories for the analysis of oestrogen receptor protein. We found that there was only a 10% variation in whether tumours were classified as receptor-positive or negative, but several-fold differences in quantitative values. Similarly, studies in our own laboratories showed a 7-fold difference in quantitative value between different parts of the same tumour.

3

Prof. Kaae

If survival times are used to compare the effect of two treatment regimes (e.g. simple drug versus multiple drugs) the treatment used after any relapse may influence the results.

This may distrub the validity of any comparison.

"Aspects of the Cox model in the analysis of survival data"

Richard Kay

Department of Probability and Statistics,
The University, Sheffield, S3 7RH, U.K.

1. Introduction

The purpose of this paper is to discuss and illustrate the use of regression models in the analysis of time to response data. These models recognise the need to model the dependence of response time T on a collection of independent variables $x_1, x_2, \ldots, x_p$. Data from the Manchester Regional Breast Study (Lythgoe, Leck and Swindell (1978)), which will be considered in detail later, provides an illustration of the area in which these methods can be usefully employed. The response event of interest in this case might be 'death', independent variables considered as having a possible effect on time to death being age of patient, tumour size, tumour site, menopausal status, stage of disease and of course treatment. Inevitably data of this type involves arbitrary right censoring. Patients enter the trial sequentially in time and at data analysis many patients who have not yet responded will produce data of this type.

2. Regression models

A convenient way of specifying the regression type models to be considered is through the hazard function. If $\lambda(t;\underset{\sim}{x})$ represents the hazard function of the response time T of an individual with independent variables $\underset{\sim}{x}$ then by definition

$$\lambda(t;\underset{\sim}{x}) = \lim_{\delta t \to 0} \frac{p(t<T\leq t+\delta t \,|\, T>t;\underset{\sim}{x})}{\delta t}$$

If the response event is death $\lambda(t;\underset{\sim}{x})$ is sometimes termed the force of mortality or failure rate.

Cox (1972) proposes a model in which

$$\lambda(t;\underset{\sim}{x}) = \lambda_o(t)e^{\underset{\sim}{\beta}'\underset{\sim}{x}} \qquad (1)$$

where $\lambda_o(t)$ is an unspecified function of t and $\underset{\sim}{\beta}$ is a pxl vector of parameters which reflect the effects of the independent variables on response time. For example if x_1 is a binary treatment indicator (0: treatment A/1: treatment B) (1) has the form

$$\lambda(t;x) = \begin{cases} \lambda_o(t)e^{\beta_2 x_2+\ldots+\beta_p x_p} & x_1=0 \\ \lambda_o(t)e^{\beta_1}e^{\beta_2 x_2+\ldots+\beta_p x_p} & x_1=1 \end{cases}$$

and β_1 is seen to 'measure' the treatment effect. A positive (negative) value for β_1 indicates that the hazard function is increased (decreased) under treatment B.

Parametric forms of the Cox model have been considered by many authors. Amongst others Feigl and Zelen (1965), Glasser (1967), Breslow (1974), Lagakos (1976) and Vaeth (1978) investigate the exponential form ($\lambda_o(t)\equiv\lambda$) while Prentice (1973), Kay (1978) and Williams (1978) also consider the Weibull case ($\lambda_o(t)=\lambda t^{\alpha}$). Farewell and Prentice (1977) incorporate these and other models in a single parametric framework. In addition Feigl and Zelen (1965), Zippin and Armitage (1966) and Greenberg, Bayard and Byar (1974) under the exponential assumption consider alternatives to $e^{\underset{\sim}{\beta}'\underset{\sim}{x}}$ in modelling the effects of the independent variables.

A generalisation of (1) proposed by Kalbfleisch (1974a) allows stratification of the individuals in the study and if $\lambda(t;j,\underset{\sim}{x})$ is the hazard function for an individual in stratum j (with independent variables vector $\underset{\sim}{x}$) then this model has the form

$$\lambda(t;j,\underset{\sim}{x}) = \lambda_{oj}(t)e^{\underset{\sim}{\beta}'\underset{\sim}{x}} \qquad (2)$$

The strata for example may be defined by the values of an independent variable violating the multiplicative assumptions of (1). Further generality is obtained on allowing components of $\underset{\sim}{\beta}$ to differ between strata. Parametric forms of (2) are considered by Kay (1977). See Holt and Prentice (1974) and Holt (1978) for use of (2) with matched pairs.

Cox (1972) indicates that independent variables included in (1) may be time dependent. Examples of their use in checking model (1) assumptions concerning the multiplicative effect of $\underset{\sim}{x}$ in the two group case are given by Cox (1972) and Kalbfleisch and McIntosh (1977). These authors consider alternatives to (1) in which the exponent for one of the groups contains an additional linear term in t (Cox) and log t (Kalbfleisch and McIntosh). Prentice (1977) uses covariates of this type to provide a means of checking the 'independence' of death and censoring. Crowley and Hu (1977) in analysing data from the Stanford Heart Transplantation Program define a time-dependent "treatment" effect

$$x(t) = \begin{cases} 0 & t<y \\ 1 & t\geq y \end{cases}$$

where y is the time from entry into the program to transplant. Tissue "mismatch" scores are included in the model in a similar way. The model of Lagakos (1976) which incorporates information on time to disease "progression" y in the analysis of advanced lung cancer is equivalent to the exponential form of (1) with x(t) as above.

3. Methods of Inference

Methods of inference in the parametric models are achieved by straightforward likelihood methods and details are given in Farewell and Prentice (1977).

For model (1) let $t_{(1)}<t_{(2)}<\ldots<t_{(k)}$ denote the k ordered response times with a corresponding rearrangement $\underset{\sim}{x}_{(1)},\underset{\sim}{x}_{(2)},\ldots,\underset{\sim}{x}_{(n)}$ of the independent variables and let $R(t) = \{j:t_j\geq t\}$ be the collection of individuals who are in the study at t-0 and have not yet responded. Given that the individuals in

$R(t_{(i)})$ are at "risk" of responding at $t_{(i)}$ and that a response occurs at $t_{(i)}$ the probability that it is on the individual as observed is

$$\frac{\lambda_o(t_{(i)})e^{\underset{\sim}{\beta}'\underset{\sim}{x}_{(i)}}}{\sum_{j\in R(t_{(i)})}\lambda_o(t_{(i)})e^{\underset{\sim}{\beta}'\underset{\sim}{x}_j}}.$$

Cox (1972) then forms a likelihood for $\underset{\sim}{\beta}$ as the product of such terms, one for each response, so that

$$L(\underset{\sim}{\beta}) = \prod_{i=1}^{k}\left\{\frac{e^{\underset{\sim}{\beta}'\underset{\sim}{x}_{(i)}}}{\sum_{j\in R(t_{(i)})}e^{\underset{\sim}{\beta}'\underset{\sim}{x}_j}}\right\} \qquad (3)$$

enabling inferences on $\underset{\sim}{\beta}$ to be made without knowledge of $\lambda_o(t)$. Cox (1975) justifies the use of this expression formally in terms of partial likelihood while Kalbfleisch and Prentice (1973) have considered its construction, in the absence of time-dependent covariates as a marginal likelihood. The stratified form (2) gives a likelihood as the product over strata of terms like (3).

Use of (2) ignores information on exact censoring times recording only the intervals in which they occur. Recovery of this information, achieved by approximating the form of $\lambda_o(t)$ between failures has been considered by Johnson and Elandt-Johnson in some unpublished work. See also Crowley (1974) for discussions in the two group case and Thompson (1977) from a grouped data standpoint.

The estimation of $\lambda_o(t)$ has been considered by Kalbfleisch and Prentice (1973) who form a suitable subdivision $b_o = 0<b_1<b_2<\ldots<b_{r-1}<b_r = \infty$ of the time scale and approximate $\lambda_o(t)$ by constants between these values. Maximum likelihood estimation of the constants is achieved conditional on the estimated $\underset{\sim}{\beta}$ values. Estimation of the survivor function $\bar{F}(t;\underset{\sim}{x}) = p(T>t;\underset{\sim}{x})$ is obtained through

$$F(t;\underset{\sim}{x}) = \exp\left\{-\underset{\sim}{\beta}' \int_0^t \underset{\sim}{x}\lambda_0(u)du\right\}.$$

Breslow (1974) and Oakes (1972) provide similar results using intervals with observed deaths defining their end-points.

Although T is assumed continuous recorded response times will inevitably involve ties. Prentice and Gloeckler (1978) review and develop appropriate generalisations. Thompsons work referenced above is also relevant here.

The distributional properties of the partial likelihood estimators $\hat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$ presented by Cox (1975) indicate that the usual large sample likelihood results apply. Tests of hypothesis concerning the values of $\underset{\sim}{\beta}$ can therefore be constructed by exploiting the asymptotic normality of $\hat{\underset{\sim}{\beta}}$. Consistent estimation of the expected value of the matrix of second partial derivatives of the log likelihoog is achieved through the substitution of $\hat{\underset{\sim}{\beta}}$ values to give estimator standard errors. Alternatively a chi-square procedure based on the log likelihood ratio can be used. The simulation work of Peace and Flora (1978) on the power of these tests suggests that the likelihood ratio approach, particularly in small samples, is preferred.

It is of interest to evaluate the amount lost in terms of efficiency of estimation when model (1) is used in place of parametric alternatives. Kalbfleisch (1974b) obtains some asymptotic results in the exponential case with p=1 and no censoring. Kay (1978) extends these results for the case p=2 and in modelling the clinical trial situation investigates the effect of "uniform" censoring on these efficiencies. These results are seen to tie in with the efficiency expressions produced by Efron (1977). Oakes (1977) has provided an alternative general approach. In cases of practical importance and with reasonably large sample sizes it seems that efficiency losses are small.

4. Example

4.1 Introduction

The data set to be reported here to illustrate the use of the statistical methods outlined is from the Manchester Regional Breast Study. An analysis of these data using largely logrank procedures (Peto and Peto (1972)) is given by Lythgoe, Leck and Swindell (1978). Interest centres around the evaluation of treatment effects while adjusting for covariates. Independent variables to be considered are

$$x_1 = \log_e \text{age}, \qquad x_2 = \begin{cases} 1 & \text{tumour size} > 2\text{cm.} \\ 0 & \text{tumour size} \leq 2\text{cm.} \end{cases}$$

$$x_3 = \begin{cases} 1 & \text{tumour site lateral} \\ 0 & \text{otherwise} \end{cases}, \qquad x_4 = \begin{cases} 1 & \text{tumour site central} \\ 0 & \text{otherwise} \end{cases}$$

$$x_5 = \begin{cases} 1 & \text{postmenopausal} < 3 \text{ years} \\ 0 & \text{otherwise} \end{cases} \qquad x_6 = \begin{cases} 1 & \text{postmenopausal} \geq 3 \text{ years} \\ 0 & \text{otherwise} \end{cases}$$

$$x_7 = \begin{cases} 1 & \text{clinical stage II} \\ 0 & \text{clinical stage I} \end{cases}$$

$$y_1 = \begin{cases} 1 & \text{local + XRT, stage I} \\ 0 & \text{local, stage I} \end{cases} \qquad y_2 = \begin{cases} 1 & \text{local + XRT, stage II} \\ 0 & \text{radical, stage II} \end{cases}$$

Measurements on the above were taken on entry into the trial. Only those patients with information on all the above quantities and aged over 25 years on entry were considered in the current analysis, the data set thus being reduced from 988 to 881 patients. Several response times were considered relevant namely time to death, time to local recurrence and time to distant recurrence all measured from time of entry into the trial. These are considered in turn below.

4.2 Survival time analysis

Of the 205 deaths, 185 were distinct. The 10 pairs of tied values were randomly broken. Table I presents the Cox model $\underset{\sim}{\beta}$ estimates with estimated standard errors.

Table I Cox model fit to survival data.

Independent variable		Estimated Coefficient	Estimated standard error
x_1	age	0.004	0.672
x_2	tumour size	0.236	0.169
x_3	tumour site	0.068	0.184
x_4	tumour site	0.015	0.218
x_5	menopausal status	0.573	0.293
x_6	menopausal status	0.323	0.283
x_7	stage	0.533	0.201
y_1	treatment (I)	-0.229	0.195
y_2	treatment (II)	0.159	0.205

The value of the log likelihood at the estimates in table I is -1222.777. Interpretation of the factor effects is achieved by inspection of the coefficients. For stage I patients the local +XRT treatment is associated with improved survival while for stage II patients the radical treatment is preferred. These treatment effects however appear non-significant. Stage and perhaps tumour size and menopausal status suggest themselves as possible prognostic factors.

Table II assesses the significance of each of the independent variable coefficients by fitting reduced models with these variables omitted in turn. The treatment effects are indeed non significant. The coefficient of the stage covariate is significantly different from zero at the 1% level with patients in stage II having significantly worse prognosis.

Estimation of the underlying hazard function $\lambda_o(t)$ for the model with all independent variables included was undertaken using the Kalbfleisch and Prentice (1973) technique with r=20, $b_j - b_{j-1} = 100$, j=1,2,...,19.

Table II Evaluation of treatment effects and prognostic factors.

Estimated coefficients of independent variables included in model.									Value of log likelihood	χ^2_1 statistic
x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1	y_2		
0.001	0.229	0.073	0.028	0.575	0.327	0.640	-	0.160	-1223.472	1.39
0.045	0.238	0.067	0.012	0.574	0.313	0.613	-0.229	-	-1223.080	0.61
-	0.236	0.068	0.015	0.574	0.324	0.533	-0.229	0.159	-1222.777	0
0.042	-	0.064	0.026	0.536	0.321	0.594	-0.219	0.164	-1223.780	2.01
-0.004	0.235	-	-0.036	0.574	0.329	0.537	-0.230	0.158	-1222.847	0.14
0.006	0.236	0.061	-	0.573	0.323	0.574	-0.229	0.159	-1222.779	0
0.404	0.210	0.071	0.005	-	0.080	0.556	-0.231	0.160	-1224.556	3.56*
0.640	0.235	0.080	0.007	0.406	-	0.544	-0.232	0.147	-1223.430	1.31
-0.088	0.321	0.091	0.028	0.618	0.354	-	-0.419	0.481	-1226.205	6.86***

*p<0.10 **p<0.05 ***p<0.01 ****p<0.001

Figure I plots estimated survivor functions (log scale) for the two stage groups for a patient aged 50 years (the mean patient age) with

$$x_2=x_3=x_4=x_6=0 \qquad x_5=1.$$

Estimated 1, 3 and 5 year survival rates obtained from these functions are also given.

4.3 Analysis of times to local and distant recurrence

Formulating the problem in the competing risk setting (Holt (1978)) with cause specific hazard functions

$$\lambda(t,j;\underset{\sim}{x}) = \lim_{\delta t \to 0} \frac{p\{t<T\leq t+\delta t,\ J=j/T>t,\underset{\sim}{x}\}}{\delta t}$$

where J=1/2 is an indicator denoting the responses "death" and "local recurrence" allows analysis of time to local recurrence treating deaths without recurrence as censorings. Interchanging local recurrence and distant recurrence provides an analysis in the latter case.

Tables III and IV present Cox model fitswith respectively local recurrence or metastases and distant recurrence as the response events. Tests of significance for the coefficients using the chi-square procedure are also given. For local recurrence age is a significant factor ($p<0.05$) with younger patients having shorter times to local recurrence. The coefficients of tumour site also approach significance ($p<0.10$) suggesting that patients with lateral and central tumours recur sooner than patients with medial tumours. The stage I treatment effect is very highly significant ($p<0.001$), radiotherapy having a clear beneficial effect on prolonging time to tumour recurrence locally.

For distance recurrence the only significant effect is tumour size ($p<0.05$). Patients with larger tumours having shorter times to distant recurrence.

Table III Model fitting for local recurrence data.

	Independent variable	Estimated coefficient	Estimated standard error of coefficient	χ^2 statistic	d.f.
x_1	age	-1.263	0.628	3.94**	1
x_2	tumour size	0.211	0.160	1.77	1
x_3	tumour site	0.418	0.202	5.35*	2
x_4		0.457	0.229		
x_5	menopausal status	0.527	0.293	4.14	2
x_6		0.466	0.273		
x_7	stage	0.177	0.191	0.84	1
y_1	treatment (I)	-0.822	0.192	19.84****	1
y_2	treatment (II)	-0.226	0.227	1.00	1

Table IV Model fitting for distance recurrence data.

	Independent variable	Estimated coefficient	Estimated standard error of coefficient	χ^2 statistic	d.f.
x_1	age	-0.806	0.747	1.14	1
x_2	tumour size	0.400	0.196	4.39**	1
x_3	tumour site	-0.126	0.202	0.39	2
x_4		-0.111	0.242		
x_5	menopausal status	0.486	0.349	3.24	2
x_6		0.532	0.321		
x_7	stage	0.171	0.243	0.49	1
y_1	treatment (I)	-0.287	0.213	1.84	1
y_2	treatment (II)	0.279	0.257	1.19	1

4.4 Use of time-dependent independent variables

A Cox model fit of the stage I survival data with assessment of significant effects is presented in table V. Menopausal status is the only significant effect ($p<0.05$).

Table V Model fitting for stage I survival data.

	Independent variable	Estimated coefficient	Estimated standard error of coefficient	χ^2 statistics	d.f.
x_1	age	-0.702	0.874	0.63	1
x_2	tumour size	0.013	0.197	0	1
x_3	tumour site	-0.214	0.232	0.84	2
x_4		-0.115	0.282		
x_5	menopausal status	1.146	0.397	7.70**	2
x_6		0.688	0.385		
y	treatment	-0.221	0.195	1.29	1

To illustrate the possible use of time-dependent independent variables these data are reanalysed in table VI with the inclusion of

$$z_1(t) = \begin{cases} 0 & t<y \\ 1 & t\geq y \end{cases}$$

where y is the time to first recurrence (in a local or distant sense) and

$$z_2(t) = \log_e(e^{x_1}+t)$$

which measures the natural logarithm of current age.

Table VI Inclusion of time-dependent covariates in analysis of stage I survival data.

Independent variable	Estimated coefficient	χ^2 statistic	d.f.
x_5, x_6 menopausal status	0.858 0.446	4.25	2
y	0.067	0.11	1
z_1 current age	-0.345	0.15	1
z_2 recurrence	2.377	130.04****	1

The use of $z_1(t)$ allows information on recurrence time to be accounted for in the modelling of survival time, while inclusion of current age investigates a possible dependence on age which gives younger patients an initial preferred survival which diminishes as time from surgery increases. In addition this later analysis only the menopausal and treatment effects are included. The coefficient of z_2 is significantly different from zero ($p \ll 0.001$) indicating that patients who recur have hazard rates which increase substantially. The menopause effects in contrast to the analysis in table V appear non significant. It may be that menopausal status effects time to death only indirectly through its effect on recurrence and inclusion of the recurrence indicator accounts for the apparent survival time dependence. Additional investigation however is needed to confirm this.

It must be stressed that the data analysis presented here is merely as an illustration of the Cox model methodology and that further work is needed on this data set before any medical conclusions can be safely drawn.

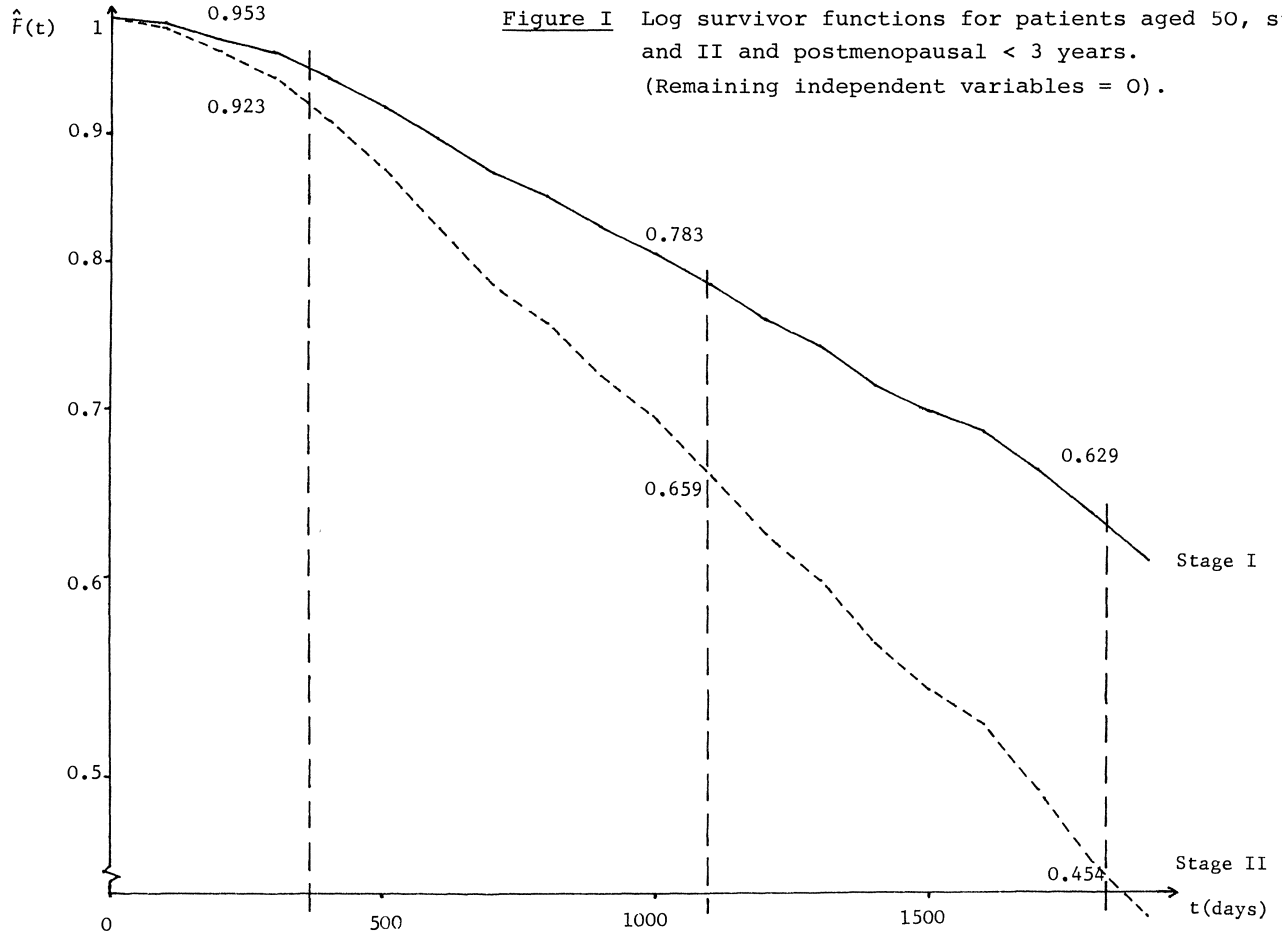


Figure I Log survivor functions for patients aged 50, stages I and II and postmenopausal < 3 years. (Remaining independent variables = 0).

References

BRESLOW, N. (1974) "Covariance Analysis of Censored Survival Data". Biometrics, 30, 89-99.

COX, D. R. (1972) "Regression models and life tables". J. R. Statist. Soc., B, 34, 187-220.

COX, D. R. (1975) "Partial likelihood". Biometrika, 62, 269-276.

CROWLEY, J. (1974) "A note on some recent likelihoods leading to the log-rank test". Biometrika, 61, 533-538.

CROWLEY, J. and HU, M. (1977) "Covariance analysis of heart transplant survival data". J. Amer. Statist. Assoc., 72, 27-36.

EFRON, B. (1977) "Efficiency of Cox's likelihood function for censored data". J. Amer. Statist. Assoc., 72, 69-75.

FAREWELL, V. T. and PRENTICE, R. L. (1977) "A study of distributional shape in life testing". Technometrics, 19, 69-75.

FEIGL, P. and ZELEN, M. (1965) "Estimation of exponential survival probabilities with concomitant information". Biometrics, 21, 826-838.

GLASSER, M. (1967) "Exponential survival with covariance". J. Amer. Statist. Assoc., 62, 561-568.

GREENBERG, R. A., BAYARD, S. and BYAR, D. (1974) "Selecting concomitant variables using a likelihood ratio step-down procedure and a method of testing goodness of fit in an exponential survival model". Biometrics, 30, 601-608.

HOLT, J. D. (1978) "Competing risk analyses with special reference to matched pair experiments". Biometrika, 65, 159-165.

HOLT, J. D. and PRENTICE, R. L. (1974) "Survival analyses in twin studies and matched pair experiments". Biometrika, 61, 17-30.

KALBFLEISCH, J. D. (1974a) "Some extensions and applications of Cox's regression and life model". Abstract. Biometrics, 30, 561.

KALBFLEISCH, J. D. (1974b) 'Some efficiency calculations for survival distributions' Biometrika, 61, 31-38.

KALBFLEISCH, J. D. and McINTOSH, A. A. (1977) "Efficiency in survival distributions with time-dependent covariates". Biometrika, 64, 47-50.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1973) "Marginal likelihoods based on Cox's regression and life model". Biometrika, 60, 267-278.

KAY, R. (1977) "Proportional Hazard Regression Models and the Analysis of Censored Survival Data". Appl. Statist., 26, 227-237.

KAY, R. (1978) "Some further asymptotic efficiency calculations for survival data regression models". Biometrika (to appear).

LAGAKOS, S. W. (1976) "A stochastic model for censored-survival data in the presence of an auxiliary variable". Biometrics, 32, 551-559.

LYTHGOE, J. P., LECK, I. and SWINDELL, R. (1978) "Manchester Regional Breast Study. Preliminary Results". Lancet, 744-747.

OAKES, D. (1972) "Contribution to discussion of paper by D. R. Cox". J. R. Statist. Soc., B, 34, 208.

OAKES, D. (1977) "The asymptotic information in censored survival data". Biometrika, 64, 441-448.

PEACE, K. E. and FLORA, R. E. (1978) "Size and power assessments of tests of hypotheses on survival parameters". J. Amer. Statist. Assoc., 73, 129-132.

PETO, R. and PETO, J. (1972) "Asymptotically efficient rank invariant test procedures". J. R. Statist. Soc., A, 135, 185-206.

PRENTICE, R. L. (1973) "Exponential survivals with censoring and explanatory variables". Biometrika, 60, 279-284.

PRENTICE, R. L. (1977). "The analysis of dose time-to-response data". Proc. of the symp. on "Dose-Response-Analysis". University of Bremen, September 1977, 66-104.

PRENTICE, R. L. and GLOECKLER, L. A. (1978) "Regression Analysis of Grouped Survival Data with Applications to Breast Cancer Data". Biometrics, 34, 57-67.

THOMPSON, W. A. (1977) "On the treatment of grouped observations in life studies". Biometrics, 33, 463-470.

VAETH, M. (1978) "Exponential lifetesting with uniform censoring. A small-sample study". Research reports No. 36, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, Denmark.

WILLIAMS, J. S. (1978) "Efficient Analysis of Weibull Survival data from Experiments on Heterogeneous Patient Populations". Biometrics, 34, 209-222.

ZIPPIN, C. and ARMITAGE, P. (1966) "Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter". Biometrics, 22, 665-672.

Comments to Dr. Kay's paper

1

N. Keiding

The fact that the hazard rate for local recurrence decreases with increasing age has to be viewed in the competing risks context, in particular since the competitor (death) has age-increasing hazard rate.

Answer to Dr. Keiding

2

Dr. Kay

Clearly age is a time varying independent variable and extensions of

the ideas of § 4.4. of my paper to incorporate this in the model in a time dependent way might provide information.

3

Dr. Crowley

The Cox model was a tremendous breakthrough: now we can do something akin to least squares regression. I would like to mention, though, that even regression with uncensored data is very difficult to do, and should be done cautiously. This is even more true with censored data, where it is more difficult to draw pictures of what is happening.

Answer to Dr. Crowley

4

Dr. Kay

I agree.

5

Dr. Haybittle

I would suggest that we do not use the term 'response time' in these analyses, as 'response' to the clinician means almost the opposite of the meaning in this context i.e. response to treatment rather than response to the hazard. 'Event time' would, I think be better.

Answer to Dr. Haybittle

6

Dr. Kay

I would like to agree with both Dr. Crowley and Dr. Haybittle.

TIME DEPENDENT COVARIATES

J.D. Kalbfleisch
Department of Statistics
University of Waterloo, Ontario, Canada

S U M M A R Y

Time dependent covariates are considered with reference to their uses and misuses in hazard rate models for the analysis of censored failure time data. Hazard models are discussed and the construction of the likelihood is reviewed with special attention being paid to the role of censoring schemes and time varying covariates. These covariates are classified as controlled, ancillary and internal covariates and some of their uses are discussed. This article is based on joint work with R.L. Prentice and R.J. MacKay and is reported on in more detail in references [4], [6] and [8].

1. I N T R O D U C T I O N

Much recent work on the analysis of the time to failure (e.g. death or recurrence) as an endpoint in a controlled clinical trial has centered around the construction of regression models for the time T to failure. Such models tend to reduce the possibility of bias in treatment comparisons by making appropriate adjustment for any imbalances which may occur with respect to auxiliary variables. In addition, these models allow the study of prognostic variables. An understanding of how such variables affect survival experience can give insight into the disease process as well as provide information useful for predicting the course of the disease.

A regression model for failure time is often most easily specified in terms of a conditional hazard function,

$$\lambda(t;\underset{\sim}{z})dt = \mathrm{pr}\{T \in [t,t+dt) \mid T \geq t,\underset{\sim}{z}\}, \qquad (1)$$

where $\underset{\sim}{z} = (z_1,\ldots,z_p)$ is a row vector of p regression variables that are measured in advance for each individual on study. We consider here two particular models of this type which will serve as the basis for examples in this paper. The general remarks, however, apply to any way of modelling (1).

Cox (1972) in his work on the proportional hazards model proposed that the covariates in (1) might be allowed to vary with time. Thus, if $x(t)$ is a single time dependent covariate and $X(t)$ is the covariate process up to time t, $\{x(u): 0 < u < t\}$, the hazard might be defined as

$$\lambda\{(t;\underset{\sim}{z},X(t)\}dt = P\{T \in [t,t+dt) \,|\, T \geq t,\underset{\sim}{z},X(t)\}.$$

In the proportional hazards model, one might take (Cox, 1972)

$$\lambda\{t;z,X(t)\}dt = \lambda_0(t)\exp\{\underset{\sim}{z}\beta + \gamma x(t)\} \qquad (2)$$

where the failure rate is affected only by the current value $x(t)$. Time dependent covariates can similarly be incorporated in other models.

2. *CLASSIFICATION OF TIME DEPENDENT COVARIATES*

Time dependent covariates may be of three distinct types and these require separate consideration.

2.1 Covariates of type I: Defined covariates. The covariate $x(t)$ may be under the control of the experimenter, or its path may be specified in advance, so that it is essentially deterministic in nature. Such covariates arise in a number of ways. For example, in a trial of long duration, it may be useful to adjust failure rates by incorporating age as a time dependent covariate. In a chemotherapy trial, cumulative dose might be taken as a covariate and provided the regimen were well specified in advance, this would be a type I time dependent covariate. On some occasions, type I covariates are introduced to check model assumptions (see section 4.1). In addition, any fixed covariate can be viewed as a time dependent covariate of this type.

To some extent, this kind of covariate is artificial since modelling could equally be done without explicitly introducing the covariate. For example, the exponential model with a type I covariate $x(t)$

$$\lambda(t;z) = \lambda \exp(z\beta + \gamma x(t))$$

can equally be modelled as (2) with $\lambda_0(t) = \lambda \exp\{\gamma x(t)\}$ and all covariates fixed.

No additional complications are presented if the covariate is under the control of the experimenter, but its level at time t is allowed to depend in some way on the previous history of the trial. For example, in testing insulation in electrical cables, a common technique is to use a step voltage test in which the voltage is increased at prespecified times until failure occures. This covariate is deterministic. There would, however, be no difficulty in allowing the voltage applied at time t to depend on the previous failure experience of other items in the trial.

2.2 Covariates of type II: Ancillary covariates. A covariate of this type is the output of a stochastic process that is external to the failure time mechanism. The value of this covariate process may affect, but is itself unaffected by, the survival experience of the study. Pollution levels related to death rates would provide an example of such a covariate. Since the marginal distribution of the covariate process is independent of the parameters in the model, it constitutes ancillary information. The conditionality principle would, therefore, suggest an analysis conditional on its observed values; this conditional approach reduces this to a type I covariate.

We let X represent the full covariate path to the cessation of observation for either deterministic or ancillary covariates. For simplicity, we exclude covariates which are conditionally deterministic. The hazard function is defined as

$$\lambda(t;z,X)dt = P\{T \in [t,t+dt) \,|\, T \geq t, z, X\}$$

and its specification is equivalent to the specification of the survivor function

$$Q(t;\underset{\sim}{z},X) = P\{T \geq t | z,X\}$$
$$= \exp\{-\int_0^t \lambda(u;z(u))du\}.$$

The estimation of the hazard function, in these cases, allows complete estimation of the survivor function given the covariate process.

Since these covariates are essentially equivalent to the fixed covariate case, we can allow $\underset{\sim}{z}$ to include both fixed covariates and the total paths of any controlled or ancillary covariates.

2.3 Covariates of type III: Internal covariates. This type of covariate arises as the output of a stochastic process specific to a study individual and so typically carries with its values information about survival or failure for that individual. An example arises in a clinical trial when some measure of a patient's general condition is made at regular intervals. Suppose at time t values of 0 and 4 are assigned to $x(t)$ for dead and no clinical evidence of disease while 3,2, and 1 represent intermediate status of increasing disability. A patient typically moves from one level to another over time and the hazard $\lambda(t;\underset{\sim}{z},x(t))$ gives the instantaneous failure rate given the current status of the covariate. The specification of $\lambda(t;\underset{\sim}{z},x(t))$ is, in this case, insufficient to specify the survivor function. Indeed,

$$Q(t;\underset{\sim}{z},X(t)) = 0 \text{ or } 1$$

depending on whether or not the item has failed by time t. Interest would often center on the marginal survivor function $P\{T \geq t|\underset{\sim}{z}\}$, but estimation of this function requires, in addition, the estimation of the stochastic process $x(t)$.

It is important to note that such covariates take values subsequent to treatment administration in a comparative trial. As a result, conditioning on the observed value $x(t)$ may mask any treatment differences that are

present. Suppose, for example, that the variate $x(t)$ as defined above is included in a comparative trial of two treatments in which, for simplicity, we assume that all individuals begin in the same state ($x(0) = 3$, say). If, for example, one treatment decelerates the passage through the levels of $x(t)$ but the death rate within each state is the same for both treatments, an analysis conditional on $x(t)$ will show no treatment difference. A marginal analysis, with the covariate $x(t)$ suppressed, however, may show that the one treatment is greatly superior. These analyses together give indirect evidence on the effect of the treatments on the process $x(t)$, but there are of course better and more direct methods of assessing this.

In the next section, the construction of the likelihood is considered for experiments of this type. The derivations reflect the comments made above.

3. THE CONSTRUCTION OF THE LIKELIHOOD

Consider n individuals to have been placed on test at time 0 and suppose that the risk of failure at time t is determined by the hazard $\lambda(t;\underset{\sim}{z},X(t))$. The failure times are assumed to be right censored and the data for the ith individual are summarized as u_i, z_i, δ_i and $X_i(u_i)$ where u_i is the failure time ($\delta_i = 1$) or the time of censoring ($\delta_i = 0$), $\underset{\sim}{z}_i$ contains the fixed information on controlled and ancillary covariates and $X_i(u_i)$ is the observed path of the internal covariate $x(t)$, $0 < t < u_i$. In this section, it is shown that, in many cases, an appropriate starting point for inference is the likelihood or partial likelihood,

$$\prod_{i=1}^{n} \lambda(u_i;z_i,X_i(u_i))^{\delta_i} \exp\{-\sum_{i=1}^{n} \int_0^{u_i} \lambda(t;\underset{\sim}{z}_i,X_i(t))dt\}. \qquad (3)$$

Conditions on the censoring mechanism and time dependent covariate $x(t)$ sufficient to make (3) appropriate are considered.

Let $H(t)$ represent the complete history of the study up to time t so that $H(t)$ records all failure and censoring information as well as complete information on all time varying covariates. Since $\{H(t)\}$ is a Markov process,

the likelihood can be constructed as a product of the conditional terms

$$\begin{aligned} pr\{H(t+dt)\,|\,H(t)\} &= pr\{D_t(dt), X_t(dt), C_t(dt)\,|\,H(t)\} \\ &= pr\{D_t(dt)\,|\,H(t)\}\,pr\{X_t(dt)\,|\,H(t), D_t(dt)\}\,pr\{C_t(dt)\,|\,H(t), D_t(dt), X_t(dt)\} \end{aligned} \tag{4}$$

where $D_t(dt)$ and $C_t(dt)$ are the sets of labels associated with individuals that have failed or are censored in $(t, t+dt)$ and $X_t(dt)$ gives the covariable information over this interval. If R_t is the set of individuals at risk at $t-0$, then

$$pr\{D_t(dt)\,|\,H(t)\} = \prod_{\ell \in D_t(dt)} \lambda(t; \underset{\sim}{z}_\ell, x_\ell(t))dt \prod_{\ell \in R_t - D_t(dt)} \{1 - \lambda(t; \underset{\sim}{z}_\ell, x_\ell(t))dt\} \tag{5}$$

where the following assumptions have been made.

1. Given $H(t)$, the failure mechanisms act independently over the interval $[t, t+dt)$.
2. For each individual in R_t, and conditional on covariates $\underset{\sim}{z}$ and $x(t)$, $pr\{\text{failure in } [t, t+dt)\,|\,H(t)\} = pr\{\text{failure in } [t, t+dt)\,|\,\text{survival to } t\}$.

It has also been assumed that the probability that a given individual in R_t is censored and also fails in $[t, t+dt)$ is $o(dt)$. Assumption 2 above is a kind of conditional independence between censoring and failure mechanisms. Censoring mechanisms which satisfy this are called independent and these include, for example, type I and type II censoring schemes, or any censoring scheme which depends only on $H(t)$ and random mechanisms external to the study. In the special case of a random censorship model where T and Y represent failure and censoring times respectively and no covariates are present with individuals being independent, it can be shown that assumption 2 is equivalent to the constant sum condition of Williams and Lagakos (1977), (see Kalbfleisch and MacKay (1978b)).

The total likelihood is a "product integral" of (4) (see Cox (1972)) where

$$L = \mathrm{pr}\{H(0)\} \prod_{o}^{T_o} [\mathrm{pr}\{H(t+dt)\,|\,H(t)\}]$$

$$= \mathrm{pr}\{H(0)\}\exp\{\lim_{\substack{\Delta t_i \to 0 \\ m \to \infty}} \sum_{i=1}^{m} \log \mathrm{pr}\{H(t_i + \Delta t_i)\,|\,H(t_i)\}$$

Here, $0 < t_0 < t_1 < \ldots < t_m = T_0$, $\Delta t_i = t_i - t_{i-1}$, and T_0 is a time after the cessation of all testing. The term in L corresponding to the failure information (5) is

$$\prod_{i=1}^{n} \lambda(u_i; \underset{\sim}{z}_i, x_i(u_i))^{\delta_i} \prod_{o}^{T_o} \prod_{\ell \in R_t - D_t(dt)} \{1 - \lambda(t; \underset{\sim}{z}_\ell, x_\ell(t))dt\}$$

which reduces to (3). The other factors are

$$\prod_{o}^{T_o} \mathrm{pr}\{X_t(dt)\,|\,H(t), D_t(dt)\} \tag{6}$$

corresponding to the instantaneous contributions of the covariate $x(t)$ and

$$\prod_{o}^{T_o} \mathrm{pr}\{C_t(dt)\,|\,H(t), D_t(dt), X_t(dt)\} \tag{7}$$

corresponding to the censoring contributions.

If either (6) or (7) depend on the parameters in the model $\lambda(t; \underset{\sim}{z}, x(t))$, the likelihood (3) is only a partial likelihood (see Cox (1975)) but can still be used for inference. If (7) depends on the parameters of interest, the censoring is called *informative* and otherwise *noninformative*. With type I and type II time dependent covariates, the factor (6) is, in the 1st case a sequence of contributions of 1 since the covariate path is deterministic given $H(t)$, and in the second case reduces to the contribution from the external process $x(t)$. If $x(t)$ is of type III, however, (6) will contain information relevant to estimation of the survivor function $P\{T > t\,|\,\underset{\sim}{z}\}$ and so (3) has a partial likelihood interpretation with respect to the estimation of this function. A complete analysis with independent censoring would require in addition the modelling of the process $x(t)$ given $\underset{\sim}{z}$ and the in-

clusion of (6) into the likelihood. An example of this approach is considered by Lagakos, Sommer and Zelen (1978) for the case where $x(t)$ is a Markov renewal process on discrete states.

4. *SOME USES OF TIME DEPENDENT COVARIATES*

In this section, some ways in which time dependent covariates can be used, are discussed.

4.1 Checks of model specifications: Cox (1972) suggested the use of time dependent covariates to check the proportional hazards specification in the model (2). In a two sample problem, we can take

$$\lambda(t;z,x(t)) = \lambda_0(t)\exp\{\beta z + \gamma x(t)\}$$

where $z = 0,1$ is a simple indicator and $x(t) = zg(t)$ where $g(t)$ is a specified function of time (e.g. $g(t) = t$ or $g(t) = \log t$). A test of $\gamma = 0$ provides a check of the proportional hazards model versus one in which the hazard ratio between the two samples varies as $ce^{\gamma g(t)}$. The covariate $x(t)$ is here being used to give a regression interpretation to a parameter γ which is, in fact, a shape parameter that differs across the two samples. Extensions of this approach to more complicated regression models provide the simplest and probably best tests available of proportional hazards.

4.2 Checks for independent censoring. The validity of the likelihood (3), which underlies most of the standard analysis of failure time data, is based on the assumption that the censoring scheme is independent. From section 3, it is evident that a censoring scheme is dependent if the probability of censoring an individual at risk at time t is related to that individual's chance of failing in a way that is not specified by $H(t)$. Thus, for example, dependent censoring occurs if the individual withdraws or is withdrawn from study when he is at high risk of failure as measured by some type III covariate $x(t)$ which is not included in the model.

If it is suspected that censoring has been allowed to depend on a type III covariate $x(t)$, a regression model for censoring with $x(t)$ as a time varying covariate can detect this (see Prentice et al (1978) and Kalbfleisch and MacKay (1978a)). The hazard of censoring may, for example, be specified as

$$\lambda(c;x(c),x(0)) = \lambda_0(c)\exp\{\beta x(0) + \gamma(x(c) - x(0))\}$$

and a test of $\gamma = 0$ provides a check. The adjustment for $x(0)$ is made since there is no difficulty in incorporating $x(0)$ as a fixed covariate. Dependence of the censoring on its level is not serious. Should the censoring scheme be found dependent, this dependence can be overcome by incorporating $x(t)$ in the model. If, however, $x(t)$ is of type III, inference is then greatly complicated and the principal quantities of interest (e.g. the marginal survivor function) may not be estimated.

As discussed in Prentice et al (1978), this general use of time dependent covariates can be extended and applied to competing risks problems in an effort to determine the extent to which different causes act independently. If a time varying indicator of the risk of death by cause i can be obtained and it is found that death by cause j is associated with this variable, there is evidence of correlated causes. This, in effect, amounts to a definition that causes are dependent if they have common time varying risk indicators and has the advantage of being verifiable unlike definitions involving latent (and necessarily unobservable) failure times.

4.3 Modelling of multivariate failure time data. Suppose that an individual being followed over time may have "failures" of two different causes. For example, in a clinical trial in cancer, patients may be observed to recurrence of disease in both primary and secondary sites. Modelling of such data is most easily done using the formulation of Cox (1972). Thus, if T_j is the time to failure of cause j, the model is specified by

$$\lambda_j(t)dt = P\{T_j \in [t,t+dt) \,|\, T_1;T_2 \geq t\},$$

$$\lambda_1(t_1|t_2)dt_1 = P\{T_1 \in [t_1,t_1+dt_1) \,|\, T_1 \geq t_1, T_2 = t_2\}$$

for $t_1 \geq t_2$, and a similar term $\lambda_2(t_2|t_1)$ for $t_2 \geq t_1$. In certain instances, one cause of failure is of particular interest in its relation to failures of the other cause. Thus, for example, comparisons of $\lambda_1(t_1)$ and $\lambda_1(t_1|t_2)$ may be primarily of interest. Additional modelling can reduce this to a problem involving time dependent covariates. Thus we might assume that

$$\lambda_1(t_1|t_2) = \lambda_1(t_1)\exp\{\beta x(t_1;t_2)\}, \quad t_1 > t_2$$

where $x(t_1;t_2)$ could take various forms to describe additional or reduced risk of a cause 1 failure given a cause 2 failure at t_2.

An interesting example of this approach is provided by Crowley and Hu (1977) in their analysis of the Stanford Heart Transplant data. In this case, T_1 represents the time to death and T_2, the waiting time to transplant. The covariate is taken to be the Heaviside function $x(t_1;t_2) = H(t_1 - t_2)$ which is 0 for $t_1 < t_2$ and 1 for $t_1 \geq t_2$.

REFERENCES

[1] Cox, D.R. (1972). Regression models and life tables (with discussion). J.R. Statist. Soc. B, 34, 187-208.

[2] Cox, D.R. (1975). Partial likelihood. Biometrika 62, 599-607.

[3] Crowley, J. and Hu, M. (1977). The covariance analysis of heart transplant data. J. Amer. Statist. Assoc., 72, 27-36.

[4] Kalbfleisch, J.D. and MacKay, R.J. (1978a). Censoring and the immutable likelihood. Submitted.

[5] Kalbfleisch, J.D. and MacKay, R.J. (1978b). On constant sum models for censored survival data. Biometrika, 65, in press.

[6] Kalbfleisch, J.D. and Prentice, R.L. (1979). The Statistical Analysis of Failure Time Data, New York: Wiley, to appear.

[7] Lagakos, S., Sommer, C.J., Zelen, M. (1978). Semi-Markov models for partially censored data. Biometrika, 65, 311-18.

[8] Prentice, R.L. and Kalbfleisch, J.D. (1978). Hazard rate models with covariates. Biometrics, 35, to appear.

[9] Prentice, R.L., Kalbfleisch, J.D., Paterson, A.V., Flournoy, N., Farewell, V.T., and Breslow, N. (1978). The analysis of failure time data in the presence of competing risks. Biometrics, 34.

[10] Williams, J.S. and Lagakos, S. (1977). Models for censored survival analysis: constant sum and variable sum models. Biometrika, 64, 215-24.

Comments to Dr. Kalbfleisch's paper

1

Dr. Wahrendorf

1.) In order to deal with time-dependent covariates it is necessary to know the value of the covariates at the actual failure times. This often is not the case and seems to make interpolation or extrapolation necessary.
2.) Cautious interpretation of time-dependent covariates should keep in mind that actually the deviation from the average value of the covariate in the risk set at the given time is studied.

Answer to Dr. Wahrendorf

2

Dr. Kalbfleisch

I think that suitable definition of the time dependent covariate can overcome Dr. Wahrendorf's first remark. For example, WBC might be measured at three monthly intervals and the covariate would be the value at the most recent measurement. I agree with his second point.

3

Dr. Prentice

I very much enjoyed your careful discussions of the likelihood function based on time-dependent covariates that are, in your terms, 'essential or internal'; that is, covariates that carry failure information. It might be worth noting that the first component of your likelihood function (concerning conditional failure information given the history of the process) can be used to produce valid inferences on factors affecting the hazard function without the necessity of modelling the sto-

chastic mechanisms giving rise to the time-varying covariates. The resulting estimation may however be inefficient if the time-dependent measurements contain most of the failure information. Also the interpretation of the regression coefficients would be useless if, for example, the censoring mechanism could not be assumed independent of the failure mechanism.

Answer to Dr. Prentice

4

Dr. Kalbfleisch

I agree with Dr. Prentice's remarks regarding the use of the first component of the likelihood for the investigation of factors affecting instantaneous failure rates. Modelling of the whole stochastic process would only be required if it were of interest to estimate features other than instantaneous failure rates such as, for example, the survivor function.

5

Dr. Haybittle

First of all, after my comment on Dr. Kay's paper, I should like to applaud Dr. Kalbfleisch's use of the term 'failure time'. I think this is even better than 'event time'.

I have been trying to think of examples in the field of breast cancer treatment where the use of time-dependent covariables would help to answer an important question. One is whether the time of local recurrence in any way influences the final survival time. Dr. Kalbfleisch made several references to the general condition of the patient as a time-dependent covariable. I doubt if there is much interest in finding out the influence of general condition throughout the course after treatment on survival. It is however important to know how long after treatment the general condition of a patient remains in a 'good' category as this might be a deciding factor in favour of one treatment even though survival times in the two arms of a trial remained the same. Presumably this question could be examined by analysing directly time to the point where patients transfer from a 'good condition' to a 'poor condition' category using a Cox type analysis without time dependent covariables.

Answer to Dr. Haybittle

6

Dr. Kalbfleisch

I think that Dr.Haybittle's analysis of the general condition example is correct. This example was chosen (in retrospect I think rather badly) since it was conceptually simple and not because it had relevance to breast cancer trials. A better choice might have been white blood count. In a chemotherapy or immunotherapy trial, one might well wish to evaluate treatment effects conditionally on the WBC.

7

Dr. Keiding

Your precise statement of how the likelihood function may be derived when all relevant internal covariation is included in the model is very important. Let me note that a martingale approach also seems feasible here. The question (also discussed by Dr. Prentice) on whether the censoring times carry important information would seem to depend on whether they are governed by the same parameters as the survival times.
Finally, do you have any way of taking care of left truncation? The construction used by DEMPSTER et al. in their paper on the EM-algorithm indicates that truncation is more difficult to handle.

Answer to Dr. Keiding

8

Dr. Kalbfleisch

The martingale approach to the lieklihood construction sounds interesting and I look forward to seeing this work. Dr. Keiding also asks about left truncation. I have not looked into the problems associated with this.

A Large Sample Study of the Estimate for the Survival Distribution in Cox'x Regression Model

Anastasios A. Tsiatis
Dept. of Statistics and Preventive Medicine
University of Wisconsin, Madison

Introduction

Regression models for survival analysis with censored observations have been used quite extensively in the past few years. One of the more widely used models is the one suggested by Cox (1972), which assumes that the hazard rate is related to a set of covariates denoted by the vector $\underset{\sim}{z} = (z_1,\ldots,z_p)$ as follows:

$$\lambda(t|\underset{\sim}{z}) = \lambda_o(t)\exp(\beta_1 z_1+\ldots+\beta_p z_p)$$

where $\lambda_o(t)$ denotes an arbitrary underlying hazard function.

The estimation and large sample properties of the regression parameters $(\beta_1,\ldots,\beta_p)$ are now well established, Cox (1972, 1975), Kalbfleisch and Prentice (1973), Breslow (1974), Liu (1978), Tsiatis (1978). This paper investigates the large sample properties of the estimate for the cumulative hazard function

$$\Lambda_o(t) = \int_0^t \lambda_o(x)dx$$

and the underlying survival distribution

$$S_o(t) = \exp - \Lambda_o(t).$$

1. Notation and Assumptions.

Let the covariate Z be a random variable with density $f(z)$.

Remark. Z is assumed to be single valued and time independent, but all the results can be extended to a vector valued set of covariates as well as time-dependent covariates.

Let Y_1, Y_2 be two positive random variables where

Y_1 denotes survival time with hazard function

$$\lambda(t|z) = \lambda_o(t)e^{\beta z},$$

and

Y_2 denotes time to censoring with arbitrary hazard function

$$\mu(t|z) = \mu(t,z).$$

The variables Y_1, Y_2 are assumed to be conditionally independent given Z.

The observable time until death or censoring will be denoted by T and the indicator for death or censoring by Δ.

$$T = \min (Y_1, Y_2)$$

$$\Delta = \begin{cases} 1 \text{ if } Y_1 \leq Y_2 & \text{(death)} \\ 0 \quad Y_1 > Y_2 & \text{(censoring).} \end{cases}$$

The observations in our study consist of n individuals associated with each is the random vector (T_i, Δ_i, Z_i), $i = 1,\ldots,n$ which are iid.

2. Some Key Relationships

The key in the estimation of the cumulative hazard function and the asymptotic properties of this estimate was in finding a relationship of the cumulative hazard as a function of other quantaties whose large sample properties can be easily established.

In so doing the following relationships are established:

The conditional probability of surviving until time x without being censored given Z is

$$H(x|z) = P(T\geq x|Z=z) = P(\min(Y_1,Y_2) \geq x|Z=z) =$$

$$P(Y_1\geq x|Z=z)P(Y_2\geq x|Z=z) = \exp\{-\int_0^x [\lambda_0(u)e^{\beta z}+\mu(u,z)]du\}.$$

The probability of surviving until time x and eventually dying given Z is

$$Q(x|z) = P(T\geq x,\Delta=1|Z=z) = \int_x^\infty \lambda_0(u)e^{\beta z}H(u|z)du.$$

Therefore the unconditional probability

$$Q(x) = P(T\geq x,\Delta=1) = \int_{-\infty}^{\infty} Q(x|z)f(z)dz$$

and the derivative is

$$dQ(x)/dx = -\lambda_0(x) \int_{-\infty}^{\infty} e^{\beta z}H(x|z)f(z)dz. \qquad (1)$$

Next we define the expectation

$$\&(e^{\beta z},x) = E[e^{\beta Z}I_{[T\geq x]}] = E[e^{\beta Z}H(x|Z)] = \int_{-\infty}^{\infty} e^{\beta z}H(x|z)f(z)dz. \qquad (2)$$

Dividing (1) by (2) we get

$$\lambda_0(x) = (-dQ(x)/dx)/\mathcal{E}(e^{\beta Z},x),$$

therefore

$$\Lambda_0(t) = \int_0^t -dQ(x)/\mathcal{E}(e^{\beta Z},x). \tag{3}$$

The terms $Q(x)$, $\mathcal{E}(e^{\beta Z},x)$ can be estimated quite naturally by the empirical probability and expectation as follows.

$$\hat{Q}(x) = (\#\text{ of deaths surviving until time } x)/n,$$

and

$$\hat{\mathcal{E}}(e^{\beta Z},x) = \sum_{i=1}^{n} e^{\beta Z_i} I_{[T_i \geq x]}/n = \sum_{j \in R(x)} e^{\beta Z_j}/n,$$

where $R(x)$ denotes the risk set at time x, or the set of indices $i = 1,\ldots,n$ corresponding to individuals who survived until time x.

An intuitive estimate of $\Lambda_0(t)$ could be obtained by substituting the empirical estimates of $Q(x)$, $\mathcal{E}(e^{\beta Z},x)$ into (3), yielding

$$\hat{\Lambda}_0(t) = \int_0^t -d\hat{Q}(x)/\hat{\mathcal{E}}(e^{\hat{\beta}Z},x) = \sum_{i \in D(t)} 1/n / \sum_{j \in R(t_i)} e^{\hat{\beta}Z_j}/n \tag{4}$$

$$= \sum_{i \in D(t)} 1 \Big/ \sum_{j \in R(t_i)} e^{\hat{\beta}Z_j},$$

where $D(t)$ denotes the set of indices of those individuals who died before t, $\hat{\beta}$ is the maximized partial likelihood estimate of β given by Cox (1972).

3. Large Sample Properties of $\sqrt{n}(\hat{\Lambda}_0(t)-\Lambda_0(t))$

Since $Q(t)$, $\hat{Q}(t)$ is a survival distribution function and its empirical estimate, much is known about the asymptotic properties of $\sqrt{n}[\hat{Q}(x)-Q(x)]$, including the fact that it converges to a mean zero Gaussian process. Using standard theory of weak convergence (see Billingsley (1968)) similar results can also be established for the quantity

$$\sqrt{n}(\hat{\mathcal{E}}(e^{\beta z},x) - \mathcal{E}(e^{\beta z},x)).$$

The random function

$$A_n(t) = \sqrt{n}\{\int_0^t -d\hat{Q}(x)/\hat{\mathcal{E}}(e^{\beta z},x) - \int_0^t -dQ(x)/\mathcal{E}(e^{\beta z},x)\}$$

being a smooth functional of $\hat{Q}$ and $\hat{\mathcal{E}}$, can be shown to converge weakly to a mean zero indendent increments Gaussian process by using techniques similar to Breslow-Crowley (1974).

This is approximately what we want except that the estimate $\hat{\Lambda}_0(t)$ is evaluated at $\hat{\beta}$ instead of β. By using a Taylor series expansion we get

$$\sqrt{n}[\hat{\Lambda}_0(t)-\Lambda_0(t)] = A_n(t) - B_n(t) + \text{2nd order terms}$$

where

$$B_n|t) = \sqrt{n}(\hat{\beta}-\beta) \int_0^t -dQ(x)\mathcal{E}(ze^{\beta z},x)/[\mathcal{E}(e^{\beta z},x)^2.$$

Asymptotic normality of $\sqrt{n}(\hat{\beta}-\beta)$ has been established by a stochastic integral approach in Tsiatis (1978), as well asymptotic independence of $\sqrt{n}(\hat{\beta}-\beta)$ and the random process $A_n|t)$.

The final result is that $\sqrt{n}[\hat{\Lambda}_0(t)-\Lambda_0(t)]$ converges weakly to a mean zero Gaussian process, say V(t), whose covariance structure is given by

$$\mathrm{Cov}(V(s),V(t)) = \int_0^s -dQ(x)/[\mathcal{E}(e^{\beta z},x)]^2 + \sigma^2(\hat{\beta}) \int_0^s -dQ(x)\mathcal{E}(ze^{\beta z},x)/[\mathcal{E}(e^{\beta z},x)]^2 \times$$

$$\int_0^t - dQ(x)\mathcal{E}(ze^{\beta z},x)/[\mathcal{E}(e^{\beta z},x)]^2,$$

where $s \leq t$, and

$$\sigma^2(\hat{\beta}) = [\int_0^\infty -dQ(x)\{\mathcal{E}(z^2e^{\beta z},x)/\mathcal{E}(e^{\beta z},x) - [\mathcal{E}(ze^{\beta z},x)/\mathcal{E}(e^{\beta z},x)]^2\}]^{-1}.$$

For practical purposes, (i.e. confidence intervals) the asymptotic variance of $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ can be estimated by substituting $\hat{Q}$, $\hat{\mathcal{E}}$ into the above formulas, yielding

$$n\{\sum_{i\in D(t)} 1/(\sum_{j\in R(t_i)} e^{\hat{\beta}Z_j})^2 + \hat{\sigma}^2(\hat{\beta})[\sum_{i\in D(t)}\{\sum_{R(t_i)} z_j e^{\hat{\beta}Z_j}/(\sum_{R(t_i)} e^{\hat{\beta}Z_j})^2\}]^2,$$

where $\hat{\sigma}^2(\hat{\beta})$ is the inverse of the estimated information matrix in Cox's likelihood.

Since $S_0(t) = \exp - \Lambda_0(t)$, we can estimate $\hat{S}_0(t)$ by $\exp - \hat{\Lambda}_0(t)$, and by a simple application of the δ-method we get

$$\hat{\mathrm{Var}}(\hat{S}_0(t)) = [\hat{S}_0(t)]^2\hat{\mathrm{Var}}(\hat{\Lambda}_0(t)).$$

The results can be extended to

(1) p-covariates.

(2) Time-dependent covariates.

(3) The estimation of the cumulative hazard function or survival distribution evaluted for particular values of the covariates.

$\Lambda(t|z_0) = \Lambda_0(t)e^{\beta z_0}$ is estimated by

$\hat{\Lambda}_0(t)e^{\hat{\beta}z_0}$ and the

variance of $\sqrt{n}[\hat{\Lambda}_0(t)e^{\hat{\beta}z_0} - \Lambda_0(t)e^{\beta z_0}]$ is

$$e^{2\beta z_0}\left[\int_0^t -dQ(x)/[\mathcal{E}(e^{\beta z},x)]^2 + \sigma^2(\hat{\beta})\left[\int_0^t -dQ(x)\{[\mathcal{E}(ze^{\beta z},x)/\mathcal{E}(e^{\beta z},x) - z_0]/\mathcal{E}(e^{\beta z},x)\}\right]^2\right]$$

A consistent estimate can be obtained by substituting the appropriate empirical estimates, $\hat{Q}$, $\hat{\mathcal{E}}$.

References.

[1] Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.

[2] Breslow, N. (1974). Covariance analysis of censored survival data. Biometrics 30 89-99.

[3] Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2 437-453.

[4] Cox, D.R. (1972). Regression models and life tables. J. Roy. Statist. Soc. B 34 187-220.

[5] Cox, D.R. (1975). Partial likelihood. Biometrika 62 269-276.

[6] Kalbfleisch, J.D. and Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika 60 267-278.

[7] Liu, P.Y. (1978). Large sample theory of the MLE based on Cox's regression model for survival data. Unpublished Ph.D. Dissertation, University of Wisconsin, Madison.

[8] Tsiatis, A.A. (1978). A large sample study of the estimate for the integrated hazard function in Cox's regression model for survival data. University of Wisconsin Technical Report No. 526.

Comments to Dr. Tsiatis's paper

1

Dr. Prentice

I have a comment and two questions:

(1) In response to an earlier question concerning the efficiency of the proposed survivor function estimator (Breslow's estimator) it may be worth noting that this estimator can be viewed as a non-parametric maximum likelihood estimator of the survivor function.

(2) Can you give an intuitive reason why the asymptotic independence of the cumulative hazard function (at the true β) and the regression estimator $\hat{\beta}$ should be uneffected by a relocation of the regression variables?

(3) Do you have a feeling for necessary conditions on the regression variable (weaker than boundedness) for the asymptotic theory to hold?

Answer to Dr. Prentice

2

Dr. Tsiatis

(1) I am in full agreement

(2) Cox's regression model assumes that the hazard function is related to a covariate z as

$\lambda_o(t)\ e^{\beta z}$.

If we relocate z, say $z' = z - k$ then the hazard function can be written as

$\lambda'_o(t)\ e^{\beta z'}$, where $\lambda'_o(t) = \lambda_o(t)\ e^{\beta k}$.

Therefore the cumulative hazard function

$\Lambda'_o(t) = \Lambda'_o(t)\ e^{\beta k}$,

and the estimate of $\Lambda_o(t)$ (at the true β) would be equal to $\hat{\Lambda}_o(t)\ e^{\beta k}$. That is the estimate of the cumulative hazard function when the covariates are relocated is just a constant times the previous estimate which would preserve asymptotic independence with the estimate $\hat{\beta}$.

(3) All the work in my technical report (where the details of my talk are contained TR No. 526) assumes that the covariates are bounded. In most applications I feel that this assumption is reasonable and to weaken this condition would prove a difficult mathematical exercise with no obvious benefits.

3

Prof. Müller

Dr. Tsiatis' estimate implicitly uses an estimate in the tails of a population. Such a procedure is known to be extremely sensitive to deviations from the theoretical model. This seems to greatly reduce its usefulness in practice.

Answer to Prof. Müller

4

Dr. Tsiatis

The estimate of the survival distribution in the lower tail when censoring tends to be heavy is indeed unstable, as is reflected by the asymptotic variance which gets large. The estimate though in the upper tail and in the middle are good and this is generally where more precise estimates are wanted. In order to estimate survival in the lower tail more precisely censoring must be light in those areas. (i.e. you cannot estimate 10 yr. survival if most individual are not observed that long.)

Also as Dr. Prentice commented these estimates are maximum likelihood, in a non-parametric sense, and are as efficient as you can get under the restricted assumptions made.

Comparing Treatments, Adjusting for Competing Risks

M. L. Moeschberger
Department of Statistics
University of Missouri
Columbia, Missouri 65211

Summary

Often some concomitant information about the experimental units in a study accompanies the usual survival time information. This information may identify the experimental units with two or more treatment groups (in which case, indicator variables are usually employed) and/or the experimental units may enter the study with certain well-defined characteristics which may need to be considered in evaluating survival functions. This paper is concerned with incorporating such information into one's analysis when there are $k(>1)$ competing causes of failure. The theoretical lifetimes associated with each cause are assumed to be independent. Two different approaches are considered - viz., i) assuming the hazard rates to be arbitrary and ii) assuming the underlying life distributions to be completely specified apart from unknown parameters. The methods discussed in this paper are treated in more detail in the monograph by David and Moeschberger (1978).

Key words: Concomitant information; competing risks; Cox-regression models

1. Introduction

A frequently encountered problem in the analysis of survival data is that of adjusting the observed survival times to account for the presence of concomitant information (sometimes referred to as covariates, independent variables, or uncontrolled explanatory variables). Since

publications dealing with this problem - when there is a single cause of failure or when the causes of failure are left unspecified - have direct relevance to solving the problem when there are competing causes of failure, a brief discussion of this work will be given.

In the classical regression context, the expectation of the response variables, conditional on knowledge of the covariates, is assumed to be some function (usually linear) of the covariates. Feigl and Zelen (1965) and Zippin and Armitage (1966) recommend a similar adjustment for an underlying exponential life distribution with mean θ. That is, they take

$$E(Y|z) = \theta + \underline{\beta}'\underline{z} \ , \tag{1}$$

where Y is the response variable, $\underline{\beta}'$ is a row vector of s regression parameters, and $\underline{z}$ is a column vector of s measured covariates for an individual. One disadvantage of this method is that the iterative estimation procedure is complicated by the additional restriction which must be placed on the parameters, namely, $\theta + \underline{\beta}'\underline{z} > 0$ for all $\underline{z}$.

Other authors have introduced the covariates via the hazard rate, as we shall briefly describe below. Cox (1972) proposes a general model in which the p.d.f. of the survival time is taken as

$$p(y|\underline{z}) = r_o(y)c(\underline{z},\underline{\beta})\exp[-R_o(y)c(\underline{z},\underline{\beta})] \ , \ y > 0 \tag{2}$$

where $\underline{\beta}$ and $\underline{z}$ are as in (1) and $c(\underline{z},\underline{\beta})$ is any function of $\underline{\beta}$ and $\underline{z}$ such that $c(\underline{z},\underline{\beta})$ is unity if the covariables are ignored, i.e., $c(\underline{z},\underline{0}) = 1$. The hazard rate is

$$\begin{aligned} r(y|\underline{z}) &= p(y|\underline{z})/\overline{P}(y|\underline{z}) \\ &= r_o(y)c(\underline{z},\underline{\beta}) \ , \quad \text{where } \overline{P}(y|\underline{z}) = \Pr[Y > y|\underline{z}] \ , \end{aligned} \tag{3}$$

and the cumulative hazard rate is

$$R(y|\underline{z}) = R_o(y)c(\underline{z},\underline{\beta}) \ ,$$

where

$$R_o(y) = \int_o^y r_o(t)dt \ . \tag{4}$$

The quantities $r_o(y)$ and $R_o(y)$ can be left arbitrary. One of the attractive features of this model is that $r_o(y)$, the hazard rate when covariables are ignored, may be estimated. The particular specialization which Cox treats in detail is $c(\underline{z},\underline{\beta}) = \exp(\underline{\beta}'\underline{z})$. In future equations in this paper conditioning on $\underline{z}$ will not be explicitly stated though it will be implied.

If $r_o(y) = 1/\theta$, i.e., the underlying life distribution is exponential with mean θ when the covariables are ignored, then Cox (1964) and Glasser (1967) introduce the covariates $\underline{z}$ via the respective functional relations

$$b(\underline{z},\underline{\beta}) = 1 + \underline{\beta}'\underline{z} \tag{5}$$

and

$$c(\underline{z},\underline{\beta}) = \exp\ (\underline{\beta}'\underline{z}) \ . \tag{6}$$

Prentice (1973) studies a Weibull model, i.e., $r_o(t) = ct^{c-1}/\theta$, with $c(\underline{z},\underline{\beta}) = \exp(\underline{\beta}'\underline{z})$. Breslow (1974) discusses and compares the models described above.

The reader is referred to the monograph by David and Moeschberger (1978) for a more detailed treatment of competing risk theory in general, and topics in the next four sections in particular.

2. Proportional hazard rates (assuming common set of regression parameters, $\underline{\beta}$).

First, let us state some terminology useful to the competing risk situation. Let C_i (i=1, ..., k) denote the mutually exclusive causes of failure (sometimes termed risks leading to failure) to which each individual is subjected. Denote the p.d.f. and c.d.f. (cumulative distribution function) of the theoretical (hypothetical) lifetime Y_i (≥ 0)

by $p_i(y_i)$ and $P_i(y_i)$, respectively. In the simultaneous presence of all k causes (risks) only the $\min_{\ell} Y_\ell$ is observable, together with the actual cause of failure. Let

$$X_i = Y_i \,|\, Y_i = \min_{\ell} Y_\ell \tag{7}$$

denote the lifetime of an individual failing from cause C_i. We shall assume in this section that the risks act independently of each other, i.e. Y_ℓ are independent.

Now, following Cox's model (2), we have for the cause of failure C_i

$$p_i(y_i) = r_{oi}(y_i)c(\underset{\sim}{z},\underset{\sim}{\beta}) \exp[-R_{oi}(y_i)c(\underset{\sim}{z},\underset{\sim}{\beta})] \tag{8}$$

where $R_{oi}(y_i) = \int_o^{y_i} r_{oi}(t)dt$,

and

$$\begin{aligned} r_i(t) &= p_i(t)/\overline{P}_i(t) \\ &= r_{oi}(t)c(\underset{\sim}{z},\underset{\sim}{\beta}) \ . \end{aligned} \tag{9}$$

We shall assume, at the outset, that $r_{oi}(t)$, the hazard rate for C_i if the covariables are ignored, may be arbitrary. Thus, the total hazard rate ignoring the covariables,

$$r_o(t) = \sum_{i=1}^{k} r_{oi}(t) \ , \tag{10}$$

will also be left arbitrary.

It will be useful to let $X_{i(j)}$ denote the time to failure of the individual with the jth longest lifetime among those individuals whose failure is attributed to cause C_i. It is natural to let $\underset{\sim}{z}_{i(j)}$ represent the covariates associated with that individual. Furthermore, let $R(x_{i(j)})$ be the set of those individuals who are at risk at time $x_{i(j)}$, i.e., the set of survivors at the time $x_{i(j)} - 0$.

Employing a conditional argument similar to that of Cox, we obtain, for a particular failure at time $x_{i(j)}$ conditional on the risk set

$R(x_{i(j)})$, the likelihood that the failure is on the individual as observed is

$$\frac{r_{oi}(x_{i(j)})c(\underline{z}_{i(j)},\underline{\beta})}{\sum_{i=1}^{k}\sum_{\ell\in R(x_{i(j)})} r_{oi}(x_{i(j)})c(\underline{z}_{\ell},\underline{\beta})} \tag{11}$$

where $\underline{z}_{\ell}$, $\ell\in R(x_{i(j)})$, denotes the covariates associated with those individuals at risk at time $x_{i(j)}$.

It is clear, upon noting that (11) may be rewritten as

$$\frac{r_{oi}(x_{i(j)})}{r_{o}(x_{i(j)})} \cdot \frac{c(\underline{z}_{i(j)},\underline{\beta})}{\sum_{\ell\in R(x_{i(j)})} c(\underline{z}_{\ell},\underline{\beta})} \tag{12}$$

that this conditional likelihood will be independent of the given failure time if and only if the hazard rates are proportional, i.e. $r_{oi}(t)/r_{oj}(t)$ is independent of t for $i \neq j$ (or equivalently, $r_{oi}(t)/r_{o}(t)$ is independent of t for all i).

Now each failure contributes a factor like (12) to the conditional likelihood, so that if we take $c(\underline{z},\underline{\beta}) = \exp(\underline{\beta}'\underline{z})$ and assume proportional failure rates, then the conditional log likelihood is

$$L(\underline{\beta}) \propto \sum_{i=1}^{k}\sum_{j=1}^{n_i} \underline{\beta}'\underline{z}_{i(j)} - \sum_{i=1}^{k}\sum_{j=1}^{n_i} \log\Big[\sum_{\ell\in R(x_{i(j)})} \exp(\underline{\beta}'\underline{z}_{\ell})\Big] , \tag{13}$$

where N_i (a random variable) individuals fail from C_i . The likelihood function (13) may be maximized with respect to $\underline{\beta}$ by some well-known iterative method.

If we assume $r_{oi}(t)/r_{o}(t)$ to be independent of t , then, as we might expect, the X_i are identically distributed. This may easily be seen as follow. It is well known that the p.d.f. of X_i in (7) is

$$f_i(x_i) = \frac{1}{\pi_i}\, r_i(x_i) \prod_{\ell=1}^{k} \overline{F}_{\ell}(x_i) \tag{14}$$

$$= \frac{1}{\pi_i} r_{oi}(x_i)c(\underline{z},\underline{\beta}) \exp\ [-R_o(x_i)c(\underline{z},\underline{\beta})] \ ,$$

where

$$R_o(x_i) = \int_o^{x_i} r_o(t)\ dt \ , \tag{15}$$

and

$$\pi_i = \Pr[Y_i = \min_{\ell} Y_\ell] \tag{16}$$

$$= \int_o^{\infty} r_{oi}(x_i)c(\underline{z},\underline{\beta}) \exp\ [-R_o(x_i)c(\underline{z},\underline{\beta})]\ dx_i \ .$$

Now (14) will be independent of the cause of failure iff $r_{oi}(t)/r_o(t)$ is independent of t (cf. Sethuraman (1965), Chiang (1968), David (1970), Lee and Thompson (1974)), in which case

$$\pi_i = r_{oi}(t)/r_o(t) \tag{17}$$

and

$$f_i(x_i) = r_o(t)c(\underline{z},\underline{\beta}) \exp\ [-R_o(x_i)c(\underline{z},\underline{\beta})] \tag{18}$$

That is to say, the observed lifetimes will be identically distributed irrespective of the cause of failure iff $r_{oi}(t)/r_o(t)$ is independent of t. Also, the density of $Z = \min_{\ell} Y_\ell$ will be identically equal to those of the X_i in (18).

For a single cause of failure, Kalbfleisch and Prentice (1973) present a rigourous argument to arrive at the marginal likelihood of $\underline{\beta}$, obtained from the marginal distribution of the ranks of the observations. For continuous data, their marginal likelihood is identical to Cox's conditional likelihood (they are different when one allows the possibility of ties). In the context of competing risks, if we assume $r_{oi}(t)/r_o(t)$ to be independent of t then the X_i will be independent and identically distributed as in (18). Now, denoting the ordered lifetimes by $t_{(1)}, \ldots, t_{(n)}$, we may write the marginal likelihood of $\underline{\beta}$ as

$$L_1(\beta) \propto \int_0^\infty \int_{t_{(1)}}^\infty \cdots \int_{t_{(n-1)}}^\infty \prod_{i=1}^{n} f_i(t_{(i)})\, dt_{(n)} \cdots dt_{(1)}. \qquad (19)$$

Therefore,

$$L_1(\beta) \propto \int_0^\infty \int_{t_{(1)}}^\infty \cdots \int_{t_{(n-1)}}^\infty \prod_{i=1}^{n} [r_o(t_{(i)}) \exp(\underline{\beta}'\underline{z}_{(i)}) \exp\{-R_o(t_{(i)}) \exp(\underline{\beta}'\underline{z}_{(i)})\} dt_{(i)}]$$

$$\propto \exp(\underline{\beta}' \sum_{i=1}^{n} \underline{z}_{(i)}) / \prod_{i=1}^{n} \sum_{\ell \in R(t_{(i)})} \exp(\underline{\beta}'\underline{z}_\ell) , \qquad (20)$$

where $\underline{z}_{(i)}$ are the covariates associated with the individual failing at time $t_{(i)}$. Thus, for the case of proportional failure rates, the marginal log likelihood reduces to (13).

If we allow censoring, the structure of the model is more complicated but upon completing an argument similar to that in Kalbfleisch and Prentice, in conjunction with equations (19) and (20), the marginal likelihood will be as in (20) with n replaced by m, where $m(<n)$ denotes the number of failures. A similar comment may be made for Cox's conditional likelihood.

Once we have maximum likelihood estimates (MLE's) of $\underline{\beta}$, denoted by $\hat{\underline{\beta}}$, from equation (13), estimates of r_{oi}, $\hat{r}_{oi}$, may be obtained as in Cox (1972) or Kalbfleisch and Prentice (1973), by regarding each of the lifetimes whose failure was due to some cause other than C_i as being censored. The MLE's of the crude, partial crude, and net probabilities of death within some interval may then be calculated [cf. Chiang, 1968].

It is of interest to note that this method, though being partially nonparametric, lacks the generality of dealing with completely arbitrary hazard functions. In fact, if the assumption of proportional failure rates is imposed then a theoretical motivation has been suggested for the underlying life distributions to follow Weibull distributions with equal

shape constants (cf. Pike (1966), Lee and Thompson (1974)). Then we are in the situation of having the underlying life distributions being completely specified apart from unknown parameters (See Section 4).

Another important situation is when one has reason to assume that the underlying life distributions belong to some specific parametric family whose failure rates are not proportional (e.g. Weibull with unequal shape constants, normal, general Makeham-Gompertz, etc.). The general parametric likelihood function is given in Section 4.

3. Arbitrary hazard rates (assuming different $\underline{\beta}_i$ for each C_i).

The experimenter, through his knowledge of the failure mechanism, may have reason to use a different set of regression parameters, $\underline{\beta}_i' = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{is})$ for each cause of failure, C_i. The methods of Cox (1972) and Kalbfleisch and Prentice (1973) become directly applicable here since one may perform k different analyses, one for each cause of failure, by treating all failures other than the one under consideration as being censoring times, i.e., lifetimes whose failure was due to $C_h (h \neq i)$ may be regarded as censored in the sense that those individuals had not yet reached their theoretical time to failure from C_i.

Clearly, the conditional (or marginal) likelihood of $\underline{\beta}_i$, which depends upon knowledge of failure times, will not employ as much information for this case since only the failure times associated with C_i will be used and thus such an analysis will be inferior in that respect. However, the assumption of proportional hazards need not be made. This may be a decided advantage, depending upon the physical situation.

4. General likelihood function for specified hazard rates (i.e., parametric approach).

If one assumes the underlying life distributions to belong to some specific parametric family, then one may write down the likelihood function applicable to various frequently arising experimental situations

(cf. Moeschberger and David, 1971). Perhaps the most useful situation to consider is one in which an individual's failure is observed only if it occurs within some specified time period, i.e. the case of Type I censoring. Let M and R (random variables such that $n = R + M$) denote the total number of failures and survivors, respectively. Suppose that M_i (also random variables) individuals fail from cause C_i, i.e. $M = \sum_{i=1}^{k} M_i$, and let X_{ij} denote the time to failure of the jth individual failing from cause $C_i (i=1, \ldots k, j = 1, \ldots, M_i)$. Then the likelihood function for Type I censoring is given by

$$L_I \propto \prod_{i=1}^{k} \prod_{j=1}^{m_i} r_i(x_{ij}) \prod_{\ell=1}^{k} \bar{P}_\ell(x_{ij}) \prod_{\ell=1}^{r} \prod_{i=1}^{k} \bar{P}_i(\gamma_{(\ell)}) \ , \ x_{ij} < \gamma_{ij} \ , \qquad (21)$$

where γ_{ij} and $\gamma_{(\ell)} (\ell = 1, \ldots, r)$ denotes the censoring times of the jth individual failing from C_i and those of the r survivors, respectively.

In conjunction with (8), the log likelihood function becomes

$$\log L_I \propto \sum_{i=1}^{k} \sum_{j=1}^{m_i} \log[r_{oi}(x_{ij}) c(\underline{z}_{ij}, \underline{\beta})] - \sum_{i=1}^{k} \sum_{j=1}^{m_i} R_o(x_{ij}) c(\underline{z}_{ij}, \underline{\beta}) \qquad (22)$$
$$- \sum_{\ell=1}^{r} R_o(\gamma_{(\ell)}) c(\underline{z}_{(\ell)}, \underline{\beta}) \ ,$$

where $\underline{z}_{ij}$ and $\underline{z}_{(\ell)} (\ell = 1, \ldots, r)$ are the covariates associated with the jth individual failing from C_i and the r survivors, respectively.

If one wishes to have different β_i for each C_i, then

$$L_I \propto \prod_{i=1}^{k} L_{Ii} \qquad (23)$$

where

$$L_{Ii} = \left[\prod_{j=1}^{m_i} r_{oi}(x_{ij}) c(\underline{z}_{ij}, \underline{\beta}_i)\right] \exp\left[-\sum_{\ell=1}^{k} \sum_{j=1}^{m_\ell} R_{oi}(x_{\ell j}) c(\underline{z}_{\ell j}, \underline{\beta}_i)\right]$$
$$\exp\left[-\sum_{\ell=1}^{r} R_{oi}(\gamma_{(\ell)}) c(\underline{z}_{(\ell)}, \underline{\beta}_i)\right] \qquad (24)$$

Thus a separate maximization may be accomplished with each cause of failure.

The asymptotic variances of the MLE's of the parameters may be obtained from the information matrix in the usual manner.

5. Comparison of treatments

We shall confine our attention to testing hypotheses of equality of two or more treatment effects. The basic method will involve a standard dummy variable procedure and will employ the usual χ^2 approximation to the negative of the logarithm of the likelihood ratio statistic.

If $\underline{z}'_\ell = (z_{\ell 1}, \ldots, z_{\ell s})$ is the row vector of covariates associated with the ℓ^{th} individual and $\underline{\beta}' = (\beta_1, \ldots, \beta_s)$ is the row vector of regression coefficients associated with the respective covariables, then let

$$z_{\ell 1} = \begin{cases} 1 & \text{if the } \ell^{th} \text{ individual receives treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$z_{\ell 2} = \begin{cases} 1 & \text{if the } \ell^{th} \text{ individual receives treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$z_{\ell,t-1} = \begin{cases} 1 & \text{if the } \ell^{th} \text{ individual receives treatment t-1} \\ 0 & \text{otherwise} , \end{cases}$$

where $s \geq t - 1$.

To test $H_o: \beta_1 = \beta_2 = \ldots = \beta_{t-1} = 0$ (the hypothesis of no treatment difference) by the likelihood ratio test, compute

$$\log L (\underline{\beta}) - \log L (\underline{\beta}_o) = -\log \Lambda ,$$

where $\underline{\beta}'_o = (0, \ldots, 0, \beta_t, \ldots, \beta_s)$, and compare the observed value of $-\log \Lambda$ to the proper fractile of a Chi-square distribution with (t-1) degrees of freedom.

6. Summary

Suppose the failure of an individual can be classified into one of k(>1) competing causes of failure. If some type of concomitant information accompanies each experimental unit prior to the observation period of the experiment and if one may assume the theoretical lifetimes associated with each cause are independent, then this paper discusses some procedures for estimating the crude, partial crude, and net probabilities in competing risk theory. These procedures essentially "adjust" each probability, according to the covariables associated with each individual.

Inferences should be more accurate using such concomitant information than would be the case if the covariables were ignored.

References

Breslow, N. (1974) Covariance analysis of censored survival data. Biometrics 30: 89-99.

Chiang, C. L. (1968). Introduction to Stochastic Processes in Biostatistics. Wiley, New York.

Cox, D. R. (1964). Some applications of exponential ordered scores. J. R. Statist. Soc. B 26: 187-220.

David, H. A. (1970). On Chiang's proportionality assumption in the theory of competing risks. Biometrics 26, 336-9.

David, H. A. and Moeschberger, M. L. (1978). The Theory of Competing Risks. Griffin's Statistical Monographs, No. 39, London.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. Biometrics 21: 826-38.

Glasser, Marvin (1967). Exponential survival with covariance. J. Amer. Statist. Ass. 62: 561-68.

Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. Biometrika 60: 267-78.

Lee, L. and Thompson, W. A. Jr. (1974). Results on failure time and pattern for the series system. SIAM, Proceedings on Reliability and Biometry: 291-302.

Moeschberger, M. L. and David, H. A. (1971). Life tests under competing causes of failure and the theory of competing risks. Biometrics 27, 909-33.

Pike, M. (1966). A suggested method of analysis of a certain class of experiments in carcinogenesis. Biometrics 22: 142-61.

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables. Biometrika 60: 279-88.

Sethuraman, J. (1965). On a characterization of the three limiting types of the extreme. Sankhya A 27, 357-64.

Zippin, C. and Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. Biometrics 22: 665-72.

Comments to Dr. Moeschberger's paper

1

Dr. Haybittls

You mentioned, I believe, an assumption that the hazard rates from the different risks behaved the same way with time. This would not be a very realistic assumption for treated cancer patients where the hazard rate from their cancer may decrease with time after treatment, while the hazard rates from other causes will increase with time because of increasing age.

2

Dr. Stewart

Adjusting for competing risks is one thing but as a clinician the problem in the interpretation of results more often relates to finding out the possible inter-relationship between time related variables. By this I mean, for example the relationship between the onset of local recurrence after full primary treatment and the subsequent development of systemic general recurrence rather than their independent relationship to survival.

Why the Present Approach to Competing Risks Should be Abandoned

Anastasios A. Tsiatis
Dept. of Statistics and Preventive Medicine
University of Wisconsin, Madison

Introduction

In the present approach to competing risks it is assumed that there corresponds to each of k causes of death a positive random variable Y_i, called the ith potential survival time, that represents the age of death of an individual in the hypothetical condition that the ith cause is the only risk of death.

One aspect of many competing risks study is to estimate the net survival probability defined as

$$H_i(t) = P(Y_i > t).$$

Let the joint distribution of $Y_1,\ldots,Y_k$ be characterized by the multiple decrement function

$$H(t_1,\ldots,t_k) = P(Y_1 > t_1,\ldots,Y_k > t_k).$$

The potential survival times are contrasted with the actual survival time, say X, $X = \min(Y_1,\ldots,Y_k)$. Survival studies are based on the empirical counterparts of the crude survival function defined as $Q_i(t)$, $i = 1,\ldots,k$

$$Q_i(t) = P(Y_i > t, \bigcap_{j \neq i} Y_j > Y_i)$$

The basic question therefore can be posed as follows; If we know or can estimate $Q_1(t),\ldots,Q_k(t)$ then can we estimate $H_i(t)$?
The answer to this question is no!! unless very restrictive and unverifiable assumptions are made.

In order to see this more clearly, we first establish the relationship between the potential survival times and the crude survival function.

1. Relationship between net and crude probabilities

For simplicity, let us consider $k = 2$ risks of death.

$$Q_1(t) = P(Y_1 \geq t, Y_2 > Y_1)$$
$$Q_1(t+h) = P(Y_1 \geq t+h, Y_2 > Y_1),$$

therefore

$$Q_1(t) - Q_1(t+h) = P(t \leq Y_1 \leq t+h, Y_2 > Y_1) \approx H(t,t) - H(t+h,t).$$

Dividing by h and taking the limit as $h \to 0$, we get

$$\frac{dQ_1(t)}{d(t)} = \frac{\partial(H(t_1,t_2)}{\partial t_1} \Bigg|_{t_1=t_2=t.} \tag{1}$$

We also note that the overall survival probability

$$P(X \geq t) = Q_1(t) + Q_2(t) = H(t,t).$$

2. Three classes of multiple decrement functions.

We shall consider three classes of multiple decrement functions, each class will be consistent with any crude survival functions $Q_1(t)$, $Q_2(t)$ (consistent in the sense that it satisfies relationship (1)). Yet, each class will give different estimates for the net survival probabilities.

The three classes of functions are as follows

Note: I denotes independence

U denotes upper class

L denotes lower class

$$H^I(t_1,t_2) = \exp(-\Lambda_1(t_1)-\Lambda_2(t_2)),$$

where

$$\Lambda_i(t) = \int_0^t \lambda_i(x)dx,$$

$\lambda_i(x)$ denotes the hazard function for the ith cause of death.

$$H^U(t_1,t_2) = \exp\{1-[1+\Lambda_1(t_1)][1+\Lambda_2(t_2)]\},$$

and

$$H^L(t_1,t_2) = \exp[-\Lambda_1(t_1)-\Lambda_2(t_2)+f(\Lambda_1(t_1))f(\Lambda_2(t_2))],$$

where

$$f(x) = \begin{cases} x & \text{if } x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}.$$

The three classes of multiple decrement functions defined above will be consistent with any given crude survival functions $Q_1(t)$, $Q_2(t)$ if we define

$$\Lambda_i^I(t) = \int_0^t dQ_i(x)/Q(x), \quad Q(x) = Q_1(x) + Q_2(x),$$

$$\Lambda_i^U(t) = \exp\{-\int_0^t dQ_i(x)/[Q(x)(1-\log Q(x))]\} - 1$$

$$\Lambda_i^L(t) = 1 - \exp\{-\int_0^t dQ_i(x)/[Q(x)(1+\log Q(x))]\} \quad \text{for } t \leq T_o$$

$$\Lambda_i^L(T_o) - \int_{T_o}^t dQ_i(x)/Q(x) \quad \text{for } t > T_o$$

where T_o is defined as $Q(T_o) = e^{-1}$.

The implication of the three different models would be that the net survival function would have the following relation;

$$H_i^U(t) > H_i^I(t) > H_i^L(t).$$

Note: Peterson (1976), has established absolute bounds for the net survival probabilities; namely

$$Q_1(t) + Q_2(t) \leq H_1(t) \leq Q_1(t) + Q_2(0).$$

An Illustrative Example

To illustrate the above results, mortality data on mice, Hoel (1972), was looked at. We considered two risks of death, reticulum cell sarcoma and all other causes of death.

The empirical crude survival probability of reticulum cell sarcoma is illustrated in Figure 1. The function is approximated by a polynomial using a least squares fit. The degree of the polynomial was increased until the fit

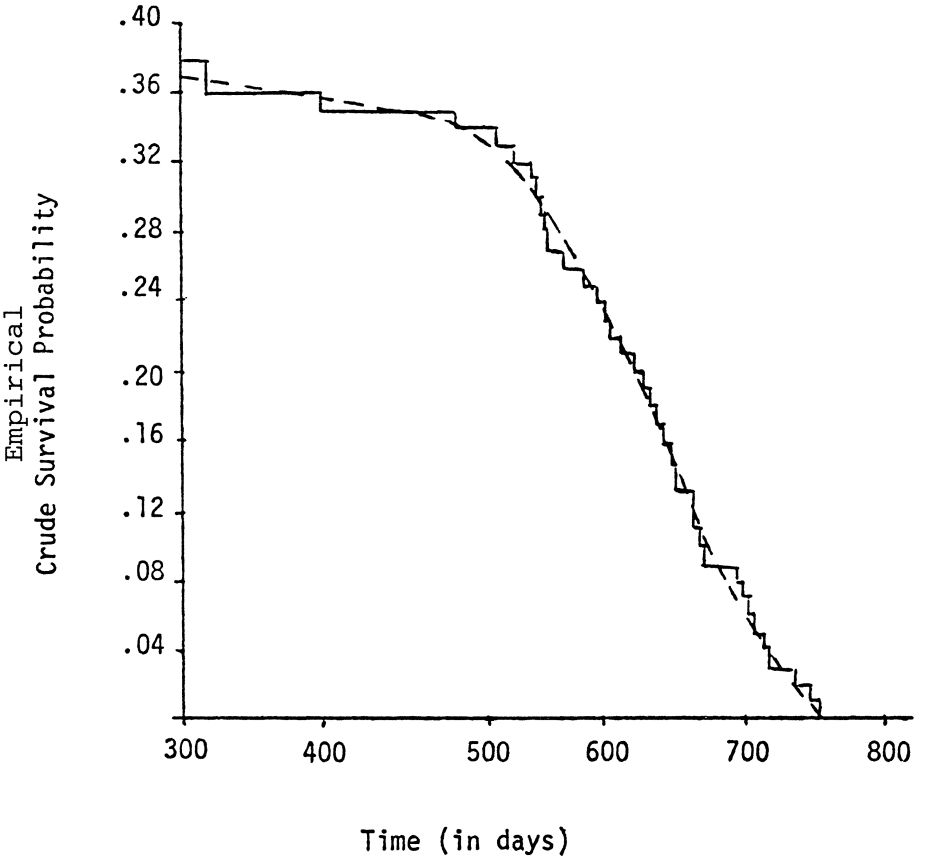

.40
.36
.32
.28
.24
.20
.16
.12
.08
.04
Empirical
Crude Survival Probability
300
400
500
600
700
800
Time (in days)

Figure 1.

was subjectively considered to be "close". (There is no special significance in the manner of fitting, we just wanted a smooth curve which corresponds closely to the crude survival function for illustrative purposes)

In Figure 2, we illustrate the three possible net survival functions of reticulum cell sarcoma resulting from the three classes of multiple decrement functions. These are also compared to the absolute Peterson bounds (dotted lines).

Concluding remarks

In the previous example, the same crude survival functions are consistent with three different multiple decrement functions resulting in very different net survival probabilities. In fact, we observe from Figure 2 that the net survival probability of reticulum cell sarcoma at 570 days can vary from 37% to 86%.

Discriminating between these results is impossible based on information available in mortality data.

When considering mortality in the presence of many causes, the estimation of net survival probabilities may prove not only misleading but also useless. The important question is not what the mortality pattern would be in the hypothetical condition where all but one cause of death is eliminated, but rather how does the overall observed mortality pattern change as a function of risk factors. This question is addressed very eloquently in the paper by Prentice.

References

[1] Hoel, D.G. (1972). A representation of mortality data by competing risks. Biometrics 28, 475-88.

[2] Peterson, A.V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proc. Nat. Acad. Sci. 73, 11-13.

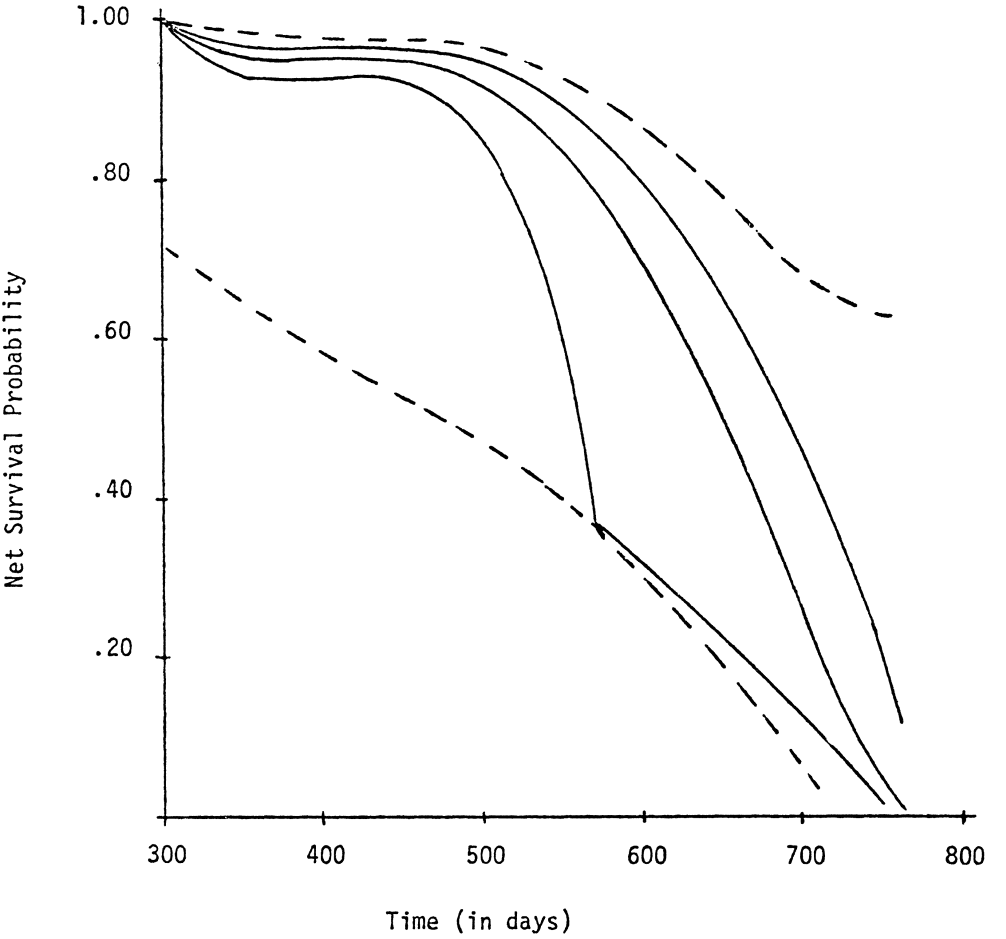

1.00
.80
.60
.40
.20
Net Survival Probability
300
400
500
600
700
800
Time (in days)

Figure 2.

Comments to Dr. Tsiatis's paper

1

Dr. Haybittle

You were very critical of age-corrected curves such as those in DUNCAN and KERR's paper, which were reproduced by Dr. v. Fournier in his talk this morning. May I explain what I think they did to produce these curves. They carried out a lifetable analysis with the data divided up into, say, yearly intervals and including deaths from all causes. They then calculated the probability of survival to different times in a normal population of the same age distribution, and used the probabilities to correct the values calculated by the lifetable analysis. Dr. Peto, would you agree that this is what they did? (Here Dr. Peto answered saying he did not agree and saying what Duncan and Kerr had done).

But if they had done as I suggested,this is a perfectly legitimate procedure for the purpose they required and can be used, by observation of the level at which the curve flattens out, to give some idea of the fraction of patients 'cured' in the sense of having the same life expectancy as the normal population. It is not a procedure very relevant to the analysis of clinical trials.

Answer to Dr. Haybittle

2

Dr. Tsiatis

I was not critical of 'age corrected curves' although I am not quite sure from Dr. Haybittle's comments exactly what he means by an age corrected curve. I was critizising the use of 'corrected survival' curves where death from other causes are treated as censored observations and life table estimates of survival are then used. This procedure estimates the exponential of minus the integral of the cause specific rates, which without the assumption of independence of the cuases of death, has no biological meaning.

3

Dr. Prentice

I would like to comment the primary reason, as I see it, for abandoning

(except perhaps in a few special cases) such notions as net and partial crude probability functions has little to do with the lack of ability to estimate such functions. The primary problem is a conceptual one: the latent failure times corresponding to specific failure types have no physical meaning so why should we be interested in the estimation of their distributions?

4

Dr. Crowley

It seems to me that your results and those of Dr. Prentice imply that the practice of plotting 'corrected' survival, were deaths from other causes are treated as censored observations, has no meaning in any biological sense and could very well be misleading. (This is in distinction to testing hypotheses, using cause-specific rates, which Dr. Prentice has shown is valid.)

Also potentially misleading are the 'relative' or 'age-corrected' survival curves, which are ratios of survival in a group to that expected from natural mortality. This does not have the interpretation of survival - just the ratio of survival curves.

Answer to Dr. Prentice and to Dr. Crowley

5

Dr. Tsiatis

I am in total agreement with both Dr. Prentice's comments and Dr. Crowley's comments.

Competing Risk Methods in Early Breast Cancer Trials

R.L. Prentice
Fred Hutchinson Cancer Research Center,
Seattle, Washington
and
Department of Biostatistics
University of Washington, Seattle

Summary

Recent work on the analysis of failure time data with competing risks is reviewed and is discussed in relation to early breast cancer trials. Central to the discussion is the modelling, estimation and interpretation of cause-specific failure rates. Possible uses of such failure rate estimators are indicated for the study of differential treatment effects on local and distant recurrence and for the identification of dependent censorship.

1. Introduction

Much of the material given in my verbal presentation at this symposium can be found in more detail in the recent paper Prentice, Kalbfleisch, Peterson, Flournoy, Farewell and Breslow (1978). The written presentation here discusses the possible application of this work to the type of early breast cancer trial to which the symposium is addressed.

Consider an 'early breast cancer' trial that aims to compare a radical form of mastectomy to some more conservative surgery, with or without post-operative irradiation (a three armed study). Data routinely recorded in such a study will include certain prestudy characteristics describing a patient's disease history as well as other personal and demographic characteristics that may be related to patient prognosis. Following entry to the study patients are followed to observe such important 'endpoints' as date and site of relapse, survival, toxicity and adherence to protocol requirements. In addition most studies will obtain additional data that attempts to assess a patient's general condition (e.g. performance status), immune competence and propensity to relapse (to the extent that suitable markers are available) on a fairly regular basis. The primary objectives of the study are, of course, (i) compare treatments in respect to the endpoints mentioned above, and (ii) learn as much as possible about the disease under study.

To be specific denote by $\underline{z} = (z_1,\ldots,z_p)$ a vector of 'covariates' defined for each study subject that will include treatment allocations and other (pre-study) prognostic factors. In the context of the study mentioned above one may, for example, define $\underline{z}$ as follows: $z_1 = 0$ or 1 according to whether the patient was allocated to conservative or radical surgery; $z_2 = 0$ or 1 according to whether or not the patient was allocated to post-operative irradiation (any two indicator variables to distinguish the three treatment groups will do); z_3 = primary tumor size (several indicator covariates could be defined corresponding to tumor size categories; $z_4 = 0$ if no positive nodes found and $z_4 = 1$ otherwise; $z_5 = 0$ if less than 4 positive nodes found and $z_5 = 1$ otherwise; $z_6 = 0$ if patient premenopausal, $z_6 = 1$ otherwise; z_7 = site within breast; $z_8 = 0$ if estrogen receptor negative and $z_8 = 1$ if estrogen receptor positive; while z_9 codes certain histological features of the disease. Other components of $\underline{z}$ may describe interactions between the treatment indicators and prognostic variables or interactions among prognostic variables themselves. For example, the inclusion of $z_{10} = z_1 \cdot z_4$ and $z_{11} = z_1 \cdot z_5$ would permit the differential effects of radical versus conservative surgery to depend on the patient's nodal status at surgery, while adjusting for the other prognostic factors listed.

Let us denote by $T \geq 0$ the time from entry into the breast cancer trial to a failure time endpoint such as death or relapse. Later the notation will be generalized to include not only the time but the type of failure. The notation will also be generalized to include covariate data recorded throughout the course of the study.

2. The Main Analyses

No attempt is made here to discuss the essential ingredients to the design of the type of breast cancer trial mentioned above. It seems clear however that large sample sizes will be required to detect treatment differences that are rather subtle on a relative basis, but may be quite important in absolute terms because of the high incidence of the disease. It is further apparent that there is usually no adequate substitute to randomized treatment allocations. Some stratification on important prognostic factors would usually be feasible and would give rise to a better controlled study. Turning to data analysis, with few exceptions all randomized patients should be included in the primary data analysis (certain violations of eligibility criteria may be exceptions). Of particular concern in data analysis is the effect of study drop-outs or losses to follow-up on the study results. This

matter will be discussed further below. A good non-technical discussion of most of these topics is given in Peto et al. (1976).

A useful framework for discussing methods for data analysis is the (instantaneous) failure rate or hazard function. The failure rate at time t for study subject with treatment and other characteristics $\underline{z}$ is defined as

$$\lambda(t;\underline{z}) = \lim_{\Delta t \to 0} P(t \leq T < t+\Delta t \mid T \geq t, \underline{z}) \; / \; \Delta t. \qquad (1)$$

Various statistical models for $\lambda(t;\underline{z})$ can be proposed in order to study the relationship between a time endpoint, T, and covariates (including treatment) $\underline{z}$. One of the most useful such models is the so-called proportional hazards model (Cox, 1972) in which covariates are presumed to affect the failure rate in a multiplicative manner; that is

$$\lambda(t;\underline{z}) = \lambda_0(t)\exp(\underline{z}\,\underline{\beta}) \qquad (2)$$

where $\underline{\beta}$ is a column p-vector of coefficients. The factor $\exp(\underline{z}\,\underline{\beta})$ is the risk of failure for an individual with characteristics $\underline{z}$ relative to that at a standard value $\underline{z} = \underline{0}$.

On the basis of (uncensored) failure times $t_1 < t_2 < \ldots < t_k$ with corresponding covariates $\underline{z}_1, \ldots, \underline{z}_k$ the relative risk parameter $\underline{\beta}$ can be conveniently estimated by minimizing the (partial) likelihood

$$L(\underline{\beta}) = \prod_{i=1}^{k} \left\{ \exp(\underline{z}_i\underline{\beta}) \; / \sum_{\ell \in R(t_i)} \exp(\underline{z}_\ell\underline{\beta}) \right\} \quad , \qquad (3)$$

where $R(t_i)$ consists of all study subjects known to be without failure just prior to failure time t_i. The model (2) is readily relaxed to allow the $\lambda_0(\cdot)$ function to differ among strata (defined from $\underline{z}$). The corresponding likelihood for $\underline{\beta}$ is simply the product of terms (3) over strata. Suitable generalizations of (3) have been proposed to accomodate tied failure times.

The methods just indicated provide the statistician with a very convenient set of tools quantifying the relationship between treatment and specific failure time endpoints while accomodating, either through modelling or stratification multiple prognostic factors. For example, under the coding scheme indicated in §1 the maximum likelihood estimator $\exp(\hat{\beta}_1)$ along with an approximate confidence interval based on the asymptotic distribution of $\hat{\beta}_1$ would provide a precise statement as to the relative risk of radical versus conservative surgery for the endpoint under examination. The method generalizes naturally to permit

such a relative risk to vary with follow-up time. Methods based on (2) and (3) also include as a special case (score test for $\underline{\beta} = \underline{0}$ when $\underline{z}$ consists only of treatment indicators) the logrank test for comparing several survival curves. The method is well suited to the examination of possible interactions between treatment differences and patient characteristics. One needs, however, to keep in mind problems related to multiple significance testing if an exploratory search is made to identify such interactions. Survival curves, censored data rank tests (e.g. logrank test and generalized Wilcoxon test) and proportional hazards analyses based on (3) or its generalizations are the principal statistical methods currently available to the statistician for the analysis of breast cancer trials or clinical studies more generally. The use of such techniques in relation to each of the study endpoints that are of the failure time variety would be expected to yield the principal study results. In view of the substantial cost and time commitment involved in carrying out a randomized clinical trial, however, it is reasonable to ask whether additional statistical methodology and analyses can provide further justification for findings arising through the use of the methods just outlined. One can also ask whether additional insight into treatment effects or natural disease history phenomena can be anticipated through a more comprehensive use of data on multiple endpoints and a more comprehensive use of explanatory data recorded over the follow-up phase of the study.

As background for discussing these topics consider the generalization of expressions (1) to (3) to include time-dependent covariate data. With the type of 'internal' covariates that involve measurements on individual study subjects the appropriate generalization of (1) to data in which a whole covariate function $Z(t) = \{\underline{z}(u): u \leq t\}$ is available on individual study subjects is as follows:

$$\lambda\{t;Z(t)\} = \lim_{\Delta t \to 0} P\{t \leq T < t+\Delta t \mid T \geq t, Z(t)\} \;/\; \Delta t \qquad (4)$$

which under a model of the type (2) can be written

$$\lambda\{t;Z(t)\} = \lambda_0(t) \exp\{\underline{z}(t)\underline{\beta}\} \; . \qquad (5)$$

A partial likelihood for $\underline{\beta}$ is again given by (3) with $\underline{z}_\ell$ replaced by $z_\ell(t_i)$ in the ith term of the product, for all subscripts ℓ.

In general expression (3) with time-dependent covariates gives a very flexible means for studying the predictive value of measurements taken over a patient's follow-up course. For example, let $z_1(u) = 0$ from

randomization up to the time of breast cancer recurrence and $z_1(u) = 1$ thereafter. Define $z_2(u) = z_1(u)\ s$, $z_3(u) = \{1-z_1(u)\}s$, where s is an indicator variable that takes value O for conservative surgery and value one for radical surgery. Examination of the corresponding coefficient $\underline{\beta}' = (\beta_1,\beta_2,\beta_3)$ allows one to compare treatments pre and post recurrence. For example, inference on β_2 contrasts mortality rates following recurrence in the two treatment groups while inference on β_3 will contrast pre recurrence mortality rates. Formal recognition not only of the time but also the type of failure can lead to further inferences along these lines.

3. Competing Risks

Suppose now that a failure time endpoint variable is relaxed to include not only the time, T, of failure but also the type of failure $J\in\{1,2,\ldots,m\}$. For example T may represent disease free survival while failure types J=1,2,3 respectively represent local recurrence, distant recurrence or death without recurrence. Alternatively T may represent time to death and values J=1,2,3 may be defined as death following a recurrence that occurred initially at a local site, death following a recurrence at a distant site or death not preceded by a recurrence.

The hazard function definition can be generalized to accomodate competing failure types as follows: Let $\lambda_j\{t;Z(t)\}$ represent the instantaneous rate of failure of type j at time t for a study subject with covariate function Z(t); that is

$$\lambda_j\{t;Z(t)\} = \lim_{\Delta t\to O} P\{t\le T<t+\Delta t,\ J=j \mid T\ge t, Z(t)\} \ / \ \Delta t, \qquad (6)$$

j=1,...,m. Three general topics that arise in the analysis of competing risk failure time data are:

(a) the association between covariates $\underline{z}$, or covariate function Z(t), and the failure rates of specific types,

(b) the interrelation among failure types,

and (c) the estimation of failure rates if some failure types were 'removed'.

Consider these topics in terms of a breast cancer trial. To be specific let T represent disease free survival while J=1,2,3 as suggested above, represent local recurrence, distant recurrence or death without disease recurrence. Topic (a) would address such questions as treatment contrasts in respect to local recurrence rates, while accomodating other prognostic factors and other failure types. Topic (b) would be concerned with such questions as whether or not patients at high risk

for local recurrence are simultaneously at high risk for distant recurrence and whether or not the strength of such association differs with treatment group. If censorship (withdrawal from study) is included as a fourth failure type (J=4) then the question of whether patients are selectively withdrawn when they are at high risk for failure (dependent censorship) also falls under topic (b). Topic (c), on the other hand, involves such questions as the effect on local recurrence rates if distant recurrences were somehow completely obviated.

Proportional hazards modelling again provides a convenient framework for addressing these topics. Suppose

$$\lambda_j\{t;Z(t)\} = \lambda_{oj}(t)\exp\{\underline{z}(t)\underline{\beta}_j\}, \quad j=1,\ldots,m, \tag{6}$$

so that both the underlying shape, $\lambda_{oj}(\cdot)$, of the cause-specific failure rate function and the relative risk parameters, $\underline{\beta}_j$, are permitted to vary arbitrarily among failure types. A partial likelihood function for $\underline{\beta}_1,\ldots,\underline{\beta}_m$ can be easily developed. It can be written

$$\prod_{j=1}^{m}\left[\prod_{i=1}^{dj}\left(\exp\{\underline{z}(t_{ji})\underline{\beta}_j\}\Big/\sum_{\ell\in R(t_{ji})}\exp\{\underline{z}_\ell(t_{ji})\underline{\beta}_j\}\right)\right] \tag{7}$$

where the failure times of type j are denoted $t_{j1},\ldots,t_{jdj}$ and $R(t_{ji})$ as usual consists of individuals at risk just prior to time t_{ji}. This likelihood is simply the product of terms of the type (3) for each failure type. It follows that testing and estimation of a particular cause-specific regression coefficients $\underline{\beta}_j$ can be carried out using the same numerical techniques as with single failure type data upon the replacement of all failure times of types other than j by censored failure times with the same value of t. The shape functions $\lambda_{oj}(\cdot)$ can be similarly estimated using single failure type procedures, at least for certain covariate types. Convenient statistical methodology then exists for studying the association between prognostic factors including treatment group and the failure rates of specific types. Suitable generalizations of (7) can be given to accomodate tied failure times of the same failure type.

Non-parametric 'survivor' function estimators are also readily generalized to competing failure types: In the above notation the overall product-limit survivor function estimator can be written

$$\hat{F}(t) = \prod_{j=1}^{m}\prod_{\{i\,|\,t_{ji}<t\}}\left[\frac{n_{ji}-d_{ji}}{n_{ji}}\right],$$

where n_{ji} and d_{ji} are the size of the risk set and the number of failure of type j at t_{ji} respectively. A non-parametric maximum likelihood estimator of $I_j(t) = P(T \leq t, J=j)$ can then be written

$$\hat{I}_j(t) = \sum_{\{i | t_{ji} < t\}} d_{ji} n_{ji}^{-1} \hat{F}(t_{ji}), \quad j=1,\ldots,m.$$

Plots, on the same figure, of $\hat{I}_j(t)$ versus t, j=1,...,m provide a useful display of competing risk data.

Consider now topic (b) concerning the interrelation among failure types. Much of the statistical literature attempts to address such questions in terms of conceptual or latent failure times. As discussed in Prentice et al. (1978) such an approach does not yield useful estimation techniques in spite of very severe assumptions. Data in addition to (T,J) or (T,J,$\underline{z}$) is required to study the relationship between different failure types. Some recent work (e.g. Wong 1977, Turnbull and Mitchell, 1978) have attempted to extrapolate back from disease prevalence data at death to draw inferences on such relationships. This approach involves supplementation of the failure type data J.

A more direct approach to problems of type (b) arises through the use of time-dependent covariates. In some situations it will be possible to define time-dependent 'risk indicators' for certain failure types which can be related to cause-specific failure rates for other failure types using (7). A time-dependent risk indicator for failure type j should reflect, perhaps crudely, the propensity of failure type j to occur on a specific individual as a function of time. Sometimes the presence of early stage disease will contribute to the risk indicator definition. In other situations less direct 'marker' data or 'risk factor' data may be all that is available.

As an example in the breast cancer setting consider a test for selective withdrawal from study. Suppose T is time from entry into the study to death or withdrawal and, for simplicity, that there are only two failure types with J=1 indicating death and J=2 indicating withdrawal. If a substantial number of patients are withdrawn from the study there is potential for serious bias in the corresponding survival curves. This is particularly important if the withdrawal rate differs between treatment groups. To take an extreme special case, it is clear that the survival data and survival comparisons would be meaningless if study subjects were selectively withdrawn from the study when thought to be moribund.

From a practical point of view one can analyze such data both under

the usual 'independent withdrawal' assumption and under a 'complete dependence' assumption in which study subjects are regarded as dead at their time of withdrawal. Consistency of treatment comparisons under the two analyses would suggest that withdrawals do not severely affect the treatment comparisons. Further understanding of the withdrawal patterns, particularly when the two analyses just indicated do not agree, can be obtained by defining one or more time-dependent risk indicators for mortality (J=1). For example performance status measurements are frequently taken at fairly regular time periods in a patient's follow-up course. Other data, such as whether or not at each time T=t the patient's breast cancer has recurred could also contribute to the definition of a risk indicator function, say $\underline{z}_1(t)$, for mortality. Define $\underline{z}(t) = (\underline{z}_1(t), \underline{z})$ where $\underline{z}$ consists of treatment and other prognostic factors. The time-dependent covariate function $Z(t) = \{\underline{z}(u): u \leq t\}$ can then be studied in relation to the withdrawal rate $\lambda_2\{t; Z(t)\}$ using (7). Components of the coefficient $\underline{\beta}_2$ that correspond to the mortality risk-indicator $\underline{z}_1(u)$ can then be examined in order to provide evidence for selective withdrawal (relative to other study subjects with the same initial characteristics at the same follow-up time). Prentice et al. (1978) give a worked example of the use of risk indicator functions in the context ot bone marrow transplantation for leukemia.

The major point to be made in respect to problem (c) is that the question of failure rates of some types given the removal of some or all other failure types is not well posed until the mechanism for cause removal is specified. This point is quite obvious in the breast cancer setting. Failure rates for distant recurrence, given that local recurrences cannot possibly happen, would be expected to depend markedly on the treatment mechanism that gave rise to the 'removal' of local recurrences. The statistical literature on such topics as 'net' and 'partial crude' probability estimation should then be avoided except in very special cases, such as situations in which different failure types arise from components of the 'system' under study that are physically or biologically independent. The question of failure rate estimation under removal of some failure types involves extrapolation from the actual study conditions in which m types of failure are operative to another set of study conditions in which only a subset are operative. Under a specific mechanism for cause removal sensible extrapolations will sometimes be possible. Again see Prentice et al. (1978) for illustration.

Acknowledgement

This work was supported by Grant Nos. CA-15704, CA-21193 and contract NO1-ES-8-2125 awarded by the National Institutes of Health, DHEW.

References

Cox, D.R. (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B 34, 187 - 220.

Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J., and Smith, P.G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part 1. Introduction and design. British Journal of Cancer 34, 585-612.

Prentice, R.L., Kalbfleisch, J.D., Peterscn, A.V., Flournoy, N., Farewell, V.T. and Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. Biometrics 34, 541-554.

Turnbull, B.W. and Mitchell, T.J. (1978). Exploratory analysis of disease prevalence from survival/sacrifice data. Biometrics 34, 555-570.

Wong, O. (1977). A competing-risk model based on the life table procedure in epidemiological studies. International Journal of Epidemiology 6, 153-159.

Comments to Dr. Prentice's paper

1

Dr. Wahrendorf

In classical regression the analysis depends on the number n of observarions and the number p of parameters. Are there at least any recommendations for Cox-model analysis available which take these parameters into account? Do time-dependent covariates in this situation count as non-time dependent ones?

Answer to Dr. Wahrendorf

2

Dr. Prentice

This question concerns the effect of sample size and the number and type of covariates on the adequacy of the asymptotic likelihood distribution theory. This is an important question which will undoubtedly receive more attention in the next few years. In general one would expect greater sample sizes to be required if the 'distribution' of covariate values is irregular (e.g. includes a small fraction of extreme and isolated values), if there are many covariated simultaneously under study and if the censoring is severe. In simple special cases such as scores test for $\underline{\beta} = 0$ (e.g. logrank test) for the comparison of several samples it is known that quite small sample sizes will suffice (e.g. 10 in each group if there is no censoring). Because of the lack of functional invariance surrounding the maximum likelihood estimator of the relative risk parameter $\underline{\beta}$, somewhat, but probably not appreciable, larger sample sizes will be required for use of the asymptotic distribution of $\hat{\underline{\beta}}$ than is the case for the corresponding score statistic. Sample sizes for use of the asymptotic distribution theory with time-dependent covariates are particularly in need of study.

3

Dr. Tsiatis

I feel this paper was the best exposition on competing risks that I have ever heard. One minor comment that I am concerned about is the analysis with the risk indicator variable. I feel that one would relate the cause specific rate of censoring to a risk indicator and if not significant they would interpret this as independence of the latent failure times of death and they would feel a sense of security in the estimation of survival probabilities using standard Kaplan-Meier estimates. This may be potentially misleading.

Answer to Dr. Tsiatis

4

Dr. Prentice
Dr. Tsiatis has made an important point that a lack of association between the failure rate function $\lambda_j\{t;Z(t)\}$ for failure type j and a 'risk-indicator' for another failure type k will only imply a lack of association between the two failure types if the risk-indicator

function is a reasonable good predictor of the type k failure risk. this could be examined by relating the time-dependent risk indicator also to the k^{th} cause-specific failure rate

$\lambda_k\{t;Z(t)\}$.

SOME EXTENSIONS OF THE LOG RANK TEST

JOHN CROWLEY

I would first like to express the appreciation all of us feel for having the chance to come to Heidelberg and participate in this conference.

1. THE LOG RANK TEST

Since I am the last speaker, if only by chance, I feel I have some obligation to try to relate a few of the concepts that have been discussed in the past few days. First let me reinforce the notion that the log rank test in its various forms can be derived from the general Cox proportional hazards model that so many of us have been discussing.

The Cox model states that the conditional risk of failing at time t, the hazard, is given by $\lambda(t,Z) = \lambda(t)\exp(Z\beta)$, where Z is a vector of prognostic factors, possibly depending on time, β is a vector of unknown coefficients, and $\lambda(t)$ the hazard for an individual with the standard set of covariates Z = 0. Dr Kalbfleisch has written the likelihood with this model as

$$L(\beta) = \prod \text{(terms delaing with failure times)} \times \prod \text{(terms dealing with time-dependent covariates)} \times \prod \text{(terms dealing with censoring times)} \quad ,$$

and has indicated that a statistic for testing the effect of prognostic

factors is given by the quadratic form

$$\left[\frac{\partial \log L(\beta)}{\partial \beta}\Big|_{\beta=0}\right]'\left[-\frac{\partial^2 \log L(\beta)}{\partial \beta^2}\Big|_{\beta=0}\right]^{-1}\left(\frac{\partial \log L(\beta)}{\partial \beta}\Big|_{\beta=0}\right).$$

In many cases this involves only the first term in the above triple product, as β does not appear in the other two terms.

1.1 The Case of Two Groups

Suppose that the hazard in the control group is given by $\lambda(t)$, and that in the experimental group $\lambda(t)\exp(\beta)$, so that the ratio of hazards, or relative risk, in the two groups is constant. This can be cast in the present framework by defining $Z = 0$ for controls, $Z = 1$ for the treatment group, and writing

$$\lambda(t,Z) = \lambda(t)\exp(Z\beta).$$

Then the rather forbidding quadratic form above reduces to a statistic which has been named by Dr Peto the log rank test and which can be expressed in the following simple way.

If the observed, distinct failure times are denoted

$$t_1 < \ldots < t_r$$

then at time t_j the observations can be cast in a 2 x 2 table:

N_{1j}	O_{1j}	$N_{1j} - O_{1j}$
N_{2j}	O_{2j}	$N_{2j} - O_{2j}$
N_j	O_j	$N_j - O_j$

where for N_{1j} is the number "at risk" in group 1, (those who have not failed and are on study at t_j) and O_{1j} is the number of failures at t_j. For group 2 N_{2j} and O_{2j} are defined similarly, and $N_j = N_{1j} + N_{2j}$, $O_j = O_{1j} + O_{2j}$. Then from standard contingency table arguments one defines the expected number of failures in group 1 at t_j as $E_j = N_{1j} \frac{O_j}{N_j}$, and the variance as

$$V_j = \frac{N_{1j} N_{2j} O_j (N_j - O_j)}{N_j^2 (N_j - 1)} \quad .$$

Then the log rank test is found by adding over failure times:

$$\frac{[\sum_{j=1}^{r} (O_j - E_j)]^2}{\sum_{j=1}^{r} V_j} \equiv \frac{(O - E)^2}{V} \quad ,$$

and referring this to tables of the Chi-Square distribution with one degree of freedom. Again, this statistic arises as a special case of the Cox proportional hazard model.

1.2 Two Groups with Strata

Suppose we wish to stratify on prognostic factors, either prospectively or retrospectively, test the difference between groups within strata, and express the results in a single statistic reflecting group differences. A possible proportional hazards model for this situation is that in stratum i, the hazard is

$\lambda_i(t)\exp(Z\beta)$, or

$\lambda_i(t)$ for controls,

$\lambda_i(t)\exp(\beta)$ for the treatment group.

Then from this stratum one can define $O_i - E_i$ and V_i as before, and the statistic resulting from the general formulation is just the sum over strata:

$$\frac{[\sum_i (O_i - E_i)]^2}{\sum_i V_i} ,$$

which is just the log rank test with stratification which Dr Peto has discussed. Note that other possible proportional hazards models could be given by $\lambda(t,Z,X) = \lambda(t)\exp(Z\beta + X\gamma)$, where $Z = 0$ and 1 for controls and treatment, X specifies the strata in some quantitative way, and γ is a vector of unknown parameters. These models lead to different statistics for testing group differences which are useful when there are a fairly large number of strata relative to the total number of patients.

Thus the "usual" log rank tests are just special cases of the Cox proportional hazards model; I hope this reinforces the notion that the log rank test can be set in a more general theoretical framework, and gives some indication that the Cox model is not an unnecessary complication but a useful extension of an old friend.

Before turning to some extensions of the log rank test, let me just mention that one can get pictorial representations of the data from the Cox model, in the form of survival curves for specific prognostic factors. As an example (perhaps not the best), one could draw two survival curves, one for controls and one for the

treatment group, with the prognostic factors fixed at their average value. Dr Tsiatis has indicated how this is done in general, and has given the necessary theory so that asymptotic standard errors for the curves can be derived.

2. THE LOG RANK TEST WITH CHANGING GROUP MEMBERSHIP

2.1 The Heart Transplant Problem

A now classical example of group membership changing with time involves the evaluation of the effect on survival of heart transplant patients. One wishes to compare the survival experience of transplanted patients to nontransplanted ones, but the only available control group consists of patients accepted into the program but who did not receive a heart (because they died first or are still waiting for a suitable donor). Early analyses made such a comparison without allowing for the fact that transplanted patients first were nontransplanted ones, but were healthy enough to wait successfully for a donor heart and receive a transplant. A possible model for this experiment is to suppose that the hazard for nontransplanted patients is $\lambda(t)$, which changes to $\lambda(t)\exp(\beta)$ at the time of transplant. This can be viewed as a Cox proportional hazards model with a time dependent covariate by defining a function $Z(t)$ which changes from 0 to 1 at the (random) time of transplant. Thus the hazard can be written as

$$\lambda(t,Z(t)) = \lambda(t)\exp(Z(t)\beta) \quad ,$$

and a test of the effect of the transplant (a test of $\beta = 0$) can be derived from the general framework given in Section 1.

The result is an extension of the log rank test in which group membership changes with time, and like the usual two group log rank

test it can be expressed simply by considering a 2 x 2 table at each death time (in either group) t_j:

N_{1j}	O_{1j}	$N_{1j} - O_{1j}$
N_{2j}	O_{2j}	$N_{2j} - O_{2j}$
N_j	O_j	$N_j - O_j$

and defining $E_j = N_{1j} \frac{O_j}{N_j}$ and

$$V_j = \frac{N_{1j} N_{2j} O_j (N_j - O_j)}{N_j^2 (N_j - 1)} .$$

The statistic is

$$\frac{[\sum_j (O_j - E_j)]^2}{\sum_j V_j} ;$$

it is again referred for testing to a Chi-Square distribution with one degree of freedom.

Here N_{1j} is the number of nontransplant patients "at risk" (alive, on study, and not transplanted) at t_j, and O_{1j} is the number of deaths in this group at t_j; N_{2j} is the number of transplanted patients at risk at t_j (alive, on study, and transplanted), and O_{2j} the number of deaths. The only difference is that patients are transferred from the first group to the second at the time of receiving their transplant; the only difficulty in the analysis is keeping track of who is at risk in each group at each death time.

The analysis outlined above can also be generalized to adjust for prognostic factors, a model for the hazard being

$$\lambda(t, Z, X) = \lambda(t)\exp(Z(t)\beta + X(t)\gamma),$$

where X(t) denotes a quantification of the relevant possibly time-dependent factors. Details of these analyses can be found in Crowley and Hu (JASA, 1977) and the references therein.

2.2 Analysis of Multiple Tumors in Rats

I have been involved recently in a series of experiments designed to test the effect of various hormonal manipulations on the induction of mammary tumors in rats. These tumors are easily palpable, so that the time to the appearance of a tumor can be measured. Also, they are easily resectable, so that tumors can be removed and the animals kept on study (to receive other tumors, get resected, etc). Time to the first tumor can be analyzed using standard survival techniques, but what can be done with multiple tumors?

One possible model is to postulate a proportional Poisson process model, with intensity parameter

$$\lambda(t,Z) = \lambda(t)\exp(Z\beta) \quad ,$$

where Z reflects the different experimental groups. Analysis via the likelihood that Dr Kalbfleisch described (Section 1) then amounts to doing a log rank test in which animals are kept in the appropriate risk sets until they die or are sacrificed, regardless of the number of tumors they develop.

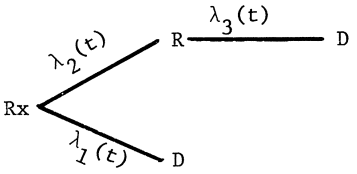
An assumption in this approach is that the risk for developing a tumor depends on time, but *not* the number of prior tumors. This assumption can be easily checked by the method of Section 2.1; within each experimental condition, define a "no tumor" group and a "tumor" group, by analogy with the nontransplant and transplant group.

A log rank test (possibly accumulated over strata, as in Section 1.2) can then be done to see if the risk for developing tumors changes with the first tumor.

In our work we found that the risk did indeed change with the first tumor, as you might imagine. We have thus been content to analyze time to first tumor, but one might consider a more complicated proportional Poisson process model, with the risk depending both on experimental condition and on the number of prior tumors.

2.3 Clinical Trials

The concept of changing group membership could also be put to use in the analysis of clinical trials in early breast cancer. Consider the following model:



where Rx represents initial therapy, R represents recurrence, and D represents death. The functions $\lambda_i(t)$ can be thought of as transition rates, completely analagous to the hazard function we have been discussing. This scheme relates to the work on stochastic models of Dr Lagakos, as well as incorporating the concept of cause-specific risks that Drs Prentice and Tsiatis have discussed. In addition, for certain purposes the time to recurrence could be though of as one of Dr Kalbfleisch's time dependent covariates.

The model suggests two points:

a) comparison of time to death for those who recurr, with time to death for those who do not, is the classical heart transplant problem which might be rephrased as a test of $\beta=0$ in the model

$\lambda_3(t) = \lambda_1(t)\exp(\beta)$: Does recurrence change the risk of dying?

b) Specific questions about differences in the cause-specific transitions between controls and a treatment group can be addressed by appropriate log rank tests. For tests of $\beta_j = 0$ in model

$$\lambda_1(t, Z) = \lambda_1(t)\exp(Z\beta_1),$$

those who recurr drop out of their respective risk sets; for

$$\lambda_2(t,Z) = \lambda_2(t)\exp(Z\beta_2),$$

those who die before recurring are removed from the risk sets; and for

$$\lambda_3(t,Z) = \lambda_3(t)\exp(Z\beta_3),$$

only those who recurr are in the risk sets.

Analysis in terms of such cause-specific transitions might allow for an increased understanding of *how* a particular treatment has its effect.

I should mention in closing this section that this model is a special case of a very general class of models, under the name "Interaction of Life History Events," being investigated by Dr Keiding (who was in attendance here earlier) and some of his Scandinavian colleagues.

3. A LOG RANK TEST FOR ORDERED ALTERNATIVES

There are certain situations which arise in survival analysis, such as comparing survival by stage for a given site of cancer, in which a statistic sensitive to a prespecified order, or trend, in the survival curves is desirable. The groups can be given a score and the Cox proportional hazard model used: assume

$$\lambda(t,Z) = \lambda(t)\exp(Z\beta) \quad ,$$

where Z gives the score for each group, then test $\beta = 0$ as in Section 1.1. However, this test will depend on the particular score given, and in that sense is somewhat arbitrary.

Consider K groups, indexed so that the desired statistic is sensitive to the alternative that group 1 has the worst survival, group 2 the next worst, etc, and group K the best. Gehan (Biometrika, 1965) suggested a Generalized Jonckheere statistic for this situation, give by $W = \sum_{i<j} W_{ij}$, where W_{ij} is the Genrealized Wilcoxon test for comparing groups i and j. A difficulty with this representation is that finding the variance is rather involved, because the terms in some are uncorrelated. Green (University of Wisconsin PhD thesis, 1979) showed that W can be rewritten

$$W = \sum_{i=1}^{k-1} W^i \quad ,$$

where $W^i = W_{i(i+1, \ldots, K)}$, the Generalized Wilcoxon statistic for comparing group i with the pooled groups i + 1 through K. This has the advantage that W^i and W^j are uncorrelated.

For the log rank test these two approaches differ and the second is also the more manageable. In a natural extension of notation, let O^i, E^i, and V^i be the observed and expected number of events, and the variance (all summed over failure times) from comparing the i^{th}

group with the pooled groups i + 1 through K. Then the proposed criterion is based on

$$\sum_{i=1}^{k-1} (O^i - E^i) \quad .$$

Since the correlation of $O^i - E^i$ and $O^j - E^j$ is 0, the statistic is

$$\left[\sum_{i=1}^{k-1} (O^i - E^i)\right]^2 \Big/ \sum_{i=1}^{k-1} v^i \quad ,$$

which can be referred for testing to a Chi-Square distribution with one degree of freedom.

I would again like to thank our hosts for their wonderful hospitality.

D i s c u s s i o n

edited by

Sarah J. Nelson

The following text is an edited version of the recorded discussion session on Thursdas 7th December 1978. The comments associated with each speaker represent only a précis of what was actually said. Dr. Haybittle, acting as chairman, began by suggesting that the discussion was structured around the questions that were sent out with the initial invitation to participants. (Appendix)

Dr. HAYBITTLE. The first two questions can be taken together. These are 'What shall we measure on whom and why?' and 'How does the clinician define the effect of a treatment?'.

Supplementary to the second question is 'How can you put such a definition into a mathematical model?'.

Let us first talk briefly about what kind of trials are worth doing in early breast cancer and what questions should next be answered. I will begin with a rather dogmatic statement about the results of trials up to now. I suggest that there are few if any questions concerning local or regional treatment which remain to be answered and therefore future trials will be connected with chemotherapy in an adjuvant form (including the possibility of hormone treatment). Perhaps somebody would like to comment on this statement.

Dr. ROBERTS. I wish I did agree with you completely as there have been many trials concerning these treatments. However, I think that the importance of local recurrence and local control have not been properly established. Although they apparently have no influence on survival their interaction with chemotherapy is not known. In trials with adjuvant therapy you must therefore be very careful about the local treatment used.

Dr. BAUM. I would like to agree with both of you. Firstly if you consider all the trials to date, without using mathematics, you can build up an impression that local therapy does not really influence outcome. Perhaps, using the methods he has described, Dr. Peto would add all these trials together and demonstrate that this is statistically sound.
The type of local therapy however is likely to be important in systemic adjuvant therapy because of its influence on tumour burden.

A more radical local therapy may lead to a more efficient systemic therapy because the initial tumour burden will be reduced. This is a potentially important interaction never before considered.

I also have a general comment to make chiefly to the mathematicians, to whom I look for advice on this matter. Our experiments are designed to test our own biological models as well as to provide you with data so you can help us describe what is happening. I am worried that we are biased and the way we record our data reflects that bias.

For example consider the recording of local recurrence as a censoring variable. This suggests local recurrence is a special factor. An alternative biological model may suggest that local recurrence is the same as distant recurrence. Perhaps we are making artificial divisions which could allow your precise mathematics to go awry.

Dr. PETO. I do not believe that is fair at all: even if the biology of the tumour cells in local and in distant recurrences is similar, there is still good reason to analyse the two events seperately. You can treat local recurrence with 6000 Rads without killing the patient. In practice this is an important difference between local and other recurrences and justifies their separate assessment, especially in trials concerned with evaluating local treatment. In such trials local recurrence is a sensitive measure of the efficiency of the treatment, even if it does not differ biologically from distant recurrence.

Returning to Dr. Haybittle's prediction that future trials will chiefly assess adjuvant therapy, another region where trials are required is in screening, where there is a great need for a randomised evaluation. To my great regret the UK government have decided not to randomise their screening trial but perhaps you could mount such a trial here in Germany. You have the mammography experience, although of course such trials are very expensive.

Dr. v. FOURNIER. We have already started such a study here two years ago. From the large studies described at this symposium it seems that improvements in survival by ten years will probably be not more than 10%. Perhaps instead of varying the treatments we should consider another way of improving the results. In our group in Heidelberg we believe early detection is likely to be a more efficient method. Mammography seems the most useful technique but we must reduce the amount of irradiation per test.

Dr. HAYBITTLE. I must remind you that less than 10% in breast cancer should not be despised as it represents many patients. I do however

accept your point concerning screening.

Dr. PETO. Surely there is still room for studies in the prevention of breast cancer as well. We know for example that early pregnancy prevents breast cancer. This is obviously not a practicable treatment, but perhaps the influences of early pregnancy on the breast could be reproduced by another method. Perhaps prevention is the easiest solution if it could be made feasible.

Dr. BAUM. Yes this is the case. All the epidemiological studies, particularly those of McMahon, have suggested the risk is related to the number of menstrual periods between menarche and first pregnancy. Roger Short, a zoologist from Edinburgh, has suggested a preventative technique using a pill which abolishes menstrual periods. This would be a feasible proposition.

Dr. HAYBITTLE. I think that we should not diverge too much into the area of prevention as we are concerned here with comparisons of treatments. However, we do seem to be agreed that a study in screening would be of value.

Dt. STEWART. Perhaps we have missed an opportunity as far as randomised screening studies are concerned. I think the majority of patients who suspect they have breast cancer would request screening if it were available. Is this the case here in Heidelberg, could you randomise patients for screening?

Dr. v. FOURNIER. Yes we can achieve this. We do not have problems like in the USA concerning the hazards of irradiation. In Germany women have a health passport at age thirty. They can have a yearly check up which includes breast palpation and we can give a mammogram if there are any findings. Up to 30% of patients are allowed to have mammography and this is a large number. Payment is made by the government via the public insurances.

Dr. HAYBITTLE. I will now go back to my original statement and summarize the comments of the meeting. As Dr. Roberts said, in a trial with adjuvant chemotherapy you must standardise the form of local therapy between the two groups to avoid the possibility of interaction. Concerning the question of what to measure, in the trials presented here similar variables have been considered in terms of survival time, time to local recurrence and time to distant recurrence.

In future, particularly with chemotherapy, perhaps we should give some measure of the'quality of life . The Karnofsky scale has been mentioned this week. Has anybody a suggestion as to how we could quantify such a difficult variable as quality of life?

Dr. PETO. Certainly the Karnofsky scale would not be applicable; it would only be useful in the really late stages of the disease.

Dr. HAYBITTLE. No, but the disadvantages of chemotherapy should be assessed in some way. Perhaps you do not feel these are very severe in the regimes used in early breast cancer at present.

Dr. BAUM. This is an important point. Suppose that chemotherapy adds a life expectancy of two years, one of these may be spent in going to the hospital and being sick each month. It is then a value judgement whether chemotherapy is desirable. There are techniques for measuring the quality of life but these are rather subjective. For example loss of earnings, days away from work and feeling sick have been suggested. More sophisticated psychological tests such as structured interviews or self rating systems are also available.

Dr. HAYBITTLE. I think the quality of life is important but I think the clinicians and related disciplines must decide how to measure it before we can attempt any analyses.

Dr. NISSEN-MEYER. In the second study of our Scandinavian group we are comparing a short term chemotherapy course with a year long one using CMF. Side effects such as the number of times vomiting and number of days feeling ill after each injection are being recorded as well as the later toxicity information. It is not clear yet how I shall analyse these but the information will be available.

Dr. ROBERTS. I would like the meeting to record its feelings concerning the value of chemotherapy. In spite of the very good results in Milan I personally regard the situation as being by no means proven. In the USA I believe chemotherapy is already given to nearly all women having a mastectomy. There is a danger that we may cause more harmful side effects than we realize if we make a judgement too quickly.

Dr. PETO. The long term survival effects of chemotherapy have not yet been established but the data show an indisputable delay in disease recurrence in premenopausal women.

Dr. ROBERTS. I would accept that effect but I think the widespread use of chemotherapy is not yet indicated.

Dr. RIBEIRO. There have been precedents where an artificial menopause was used and a delay in local recurrence found (Nissen-Meyer, Paterson and Russell). If you analysed the data too early you might have found an improvement in survival but this was in fact not the case. If there is still an imporvement at ten years we will be able to make a decision. Factors such as which treatment is easiest, which has less side effects and which is most acceptable to the patient will also have to be considered.

Dr. HAYBITTLE. We have no American clinician here but can anybody confirm the situation over there?

Dr. PRENTICE. I can only give my impression but this is consistent with what Dr. Roberts said about chemotherapy being routine. Also I agree that a delay in local recurrence has been indicated by several trials. Perhaps a clear difference in survival should not be expected as in many trials the treatment at relapse has been kept optional. In that case any suitable active agent would eventually be used, complicating the result.

Dr. STEWART. I interpret the data as indicating chemotherapy should be used but only in randomly controlled ways. When the long term effects are established it may be accepted as a universal treatment.

Dr. RIBEIRO. Yes, the results of trials are still being analysed and to begin a Phase III trial assuming the situation proved would be premature.

Dr. BAUM. I have paid several recent visits to the USA and must explain that it is almost impossible to withold adjuvant therapy for patients with early breast cancer and positive nodal histology because of the fear of malpractise litigation. Thus in virtually all trials the control arm has some chemotherapy and most patients treated outside of protocols also have chemotherapy. Yet, Dr. Fisher and Dr. Bonadonna themselves recently stated that world strategy should not be influenced by the results of two experiments.

Dr. HAYBITTLE. That sentiment leads us to the next question, 'Are the results of statistical tests convincing in such a way that the

clinician is obliged to apply a particular treatment only?'.

I think it is the context or design of the trial which is sometimes unconvincing rather than the statistical test. The result of a trial of course does not necessarily tell a clinician how to treat a particular patient.

Dr. STEWART. I would like to put forward the following view based on trials to date. I think we have evedence that patients premenopausal at presentation benefit from adjuvant chemotherapy treatments. Future trials should thus divide patients according to menopausal status and even the control arm of the premenopausal group should have chemotherapy. For premenopausal patients the question is which adjuvant scheme should be used and for postmenopausal patients whether to use one or not.

Dr. RIBEIRO. This is how we designed our Tamoxifen trial. We knew an artificial menopause worked so we felt we had to give it. There is still a danger of prejudging the situation however.

Dr. v. FOURNIER. I do not believe that the short term benefits demonstrated are sufficient to ignore the possible long term risks. You must have a control arm even for the premenopausal women.

Dr. STEWART. We already have measurements on such a control arm. The long term effect will be demonstrated by follow-up of these patients.

Dr. RIBEIRO. The difference in survival now is 20% at four years, if at ten years this is reduced to a 10% difference that is still worthwhile, the best achieved so far.

Prof. IMMICH. This example shows that even the best designed randomised study and statistical techniques fail to convince some physicians. In planning and designing studies this should be taken into account.

Dr. HAYBITTLE. The statistical tests on the data available are convincing but long term results may negate this advantage.

Dr. TSIATIS. I feel the interaction between the statistician and the clinician is important. The relationship should be one involving fre-

quent discussion on the questions and problems arising in trials.

Dr. PRENTICE. For my own information I must ask Dr. Stewart a question. It concerns the on-going study which you thought would resolve the late effects of chemotherapy. Will the control group there be eligible for late chemotherapy?

Dr. STEWART. Yes, as we saw from Dr. Valagussa's report, even those in the treatment arm have gone on to have further chemotherapy for their disease. Both groups will have this.

Dr. PETO. Dr. Prentice should not worry unduly about trials being diluted by giving chemotherapy to the controls who relapse as it does not really work. Once you have metastic disease the chemotherapy will only provide a short respite. It is not analagous, for example, to the case of relapses in acute lymphoblastic leukemia where you can give a really long remission.

Dr. PRENTICE. It may be possible that the group with chemotherapy is subject to a poorer response to late chemotherapy. This could be a problem in analysing the late effects.

Dt. STEWART. I do not think this will be significant if you consider the length of time between initial treatment and recurrent disease.

Dr. ROBERTS. The patients in the trial we are discussing had a radical mastectomy and no radiotherapy. Would patients with another form of mastectomy and quadrate radiotherapy show the same recurrence data in terms of chemotherapy? I know of no trials designed to investigate this aspect.

Dr. NISSEN-MEYER.* In the Scandinavian adjuvant chemotherapy study postoperative radiotherapy was given routinely. I have sorted out the 508 patients who had the same modified radical mastectomy and divided them into the 261 who postoperatively received cobalt irradiation to the lymph node regions, but no irradiation of the scar region, and the 247 who postoperatively received conventional X-ray treatment, including the scar region.

---

* This reply to Dr. Roberts's question was submitted after the symposium but is included here at the author's request

In both these two subseries 50% were in the chemotherapy group and 50% in the control group. The differences in total failure rates increased per year until about 10% in both subseries, in favour of the chemotherapy group. The number of local recurrences as first event in the chemotherapy groups and the control groups respectively were as follows: in the cobalt series 15 and 32, in the X-ray series 3 and 10. It seems obvious that the adjuvant chemotherapy had a preventing effect also on the local recurrences.

Dr. KAAE. I do not believe that it is proved that the benefits of a long chemotherapy course outweigh the complications introduced. In this case I think you should have a randomisation with no adjuvant therapy in premenopausal women.

Dr. BAUM. I agree with Dr. Kaae. I am involved in running trials with control arms in both pre- and postmenopausal women. We should not dismember current trials because of the results of two experiments. In addition, I think that the Milan experience could not be repeated in Britain as it involves a technically difficult and time consuming regime. If we applied CMF throughout Britain we could kill more women than we cured.

Dr. PETO. Perhaps there is a basis for compromise in the problem. The area of uncertainty is at what level of severity of the disease does the doctor feel obliged to give chemotherapy. Physicians will have different ideas about the average cut-off point but each will have a region where they are not sure whether to give it or not. If we could randomise those patients of a physician who lie within his particular region of uncertainty then we could obtain useful information. The regions of uncertainty will obviously be different for different physicians, but this is easy to cope with statistically.

Dr. LAGAKOS. I agree that we should take the opportunity to perform tests while uncertainty exists. In future years people may believe that they know all the answers and thus may not want to do tests.

Dr. SCHEURLEN. I have not found a clear formulation of the hypotheses in any of the studies which I have examined. We must not start new trials unless these hypotheses are clear. Otherwise the results may easily be contested.

Dr. STEWART. This must be the fault of the individuals doing the reporting. In most studies the hypotheses are available in the protocol and certainly have been formulated by discussion before the study begins.

Dr. HAYBITTLE. Yes, but perhaps by the time the results are reported the hypotheses are not clearly reproduced.

Dr. KAAE. The purpose in a randomised study is to show that the patient has a better life, a longer recurrence-free survival, lower frequency of local recurrence, fewer distant metastases and as few side effects as possible. The benefits of a treatment are compared with the complications in both the short and long terms.

Dr. BAUM. I agree with Dr. Scheurlen that you must define your hypotheses clearly in the questions you ask. If this is done you learn more about the disease as well as investigating treatment strategy. Such information about the disease can be built into a new hypothesis and tested as a new treatment strategy. It is not just important that one group of patients did better than another, it is also of interest to know why this happened.

Dr. LAGAKOS. Breast cancer is unique in that trial take a long time for total evaluation. In planning these trials one should therefore consider the timescales involved. Perhaps the availability of results from other studies can be anticipated and be built into the strategy. There is sometimes a value in duplicating studies but, particularly in breast cancer, it might be better to do something that, when combined with another study, produces a more significant result.

Dr. HAYBITTLE. Thank you. Let us now move onto the fourth question which is, 'How to handle departures from the process of randomisation?' Dr. Peto discussed this in his paper and Dr. Kaae demonstrated what it can involve practically. We are agreed I believe that no matter what other analyses are done an analysis which includes all patients must be performed.

Dr. LAGAKOS. I should add that the optimal solution is to avoid such departures.

Dr. HAYBITTLE. Dr. Kaae's example was typical of where you cannot do so. These examples are common.

Dr. STEWART. Do we exclude all patients who do not fulfil the criteria of eligibility and have been incorrectly entered in the trial?

Dr. HAYBITTLE. Yes if you state you are going to do so clearly at the beginning.

Dr. PETO. It is a good idea to define when you design an experiment which criteria of eligibility will lead to later exclusion if violated. In that way there can be no doubt whether to exclude a patient or not.

Dr. PRENTICE. You must be careful about making an exclusion on the basis of some condition that might be more likely to be detected in one treatment group than in another.

Dr. HAYBITTLE. We are all agreed so let us take the next question. This is concerned with statistical methods, 'In what situations should one use the logrank test, the rank sum test, Cox's model and competing-risk models?'. There are two aspects to this, one for the statisticians and the other related to how much the clinicians should be concerned with the problem. Dr. Peto has given me a dogmatic statement which I would like to read to start the discussion.

'Statistical research has now reached the point where we have optimal methods for the nonsequential analysis of clinical trials which will suffice for most situations. Further progress in the area of statistical methodology in nonsequential trials analysis cannot be expected and clinicians involved in trials only need to familiarise themselves with these concepts. These are the ideas of life tables, an O- and E-comparison in something like a logrank test and in retrospective stratification. Time period subdivision should perhaps also be included.'

Dr. TSIATIS. I do not agree at all. As clinical trials progress new problems will arise requiring new statistical methods. For example the risk indicator variable is a new and important concept.

Dr. PETO. That is concerned with subject-dependent censoring and it is not going to be relevant in the randomised clinical trial.

Dr. TSIATIS. You are assuming that nothing is ever going to change. New measures other than survival may be required, combinations of effects may be studied. You should not simplify to a point where you

begin to lose information rather than gain it.

Dr. HAYBITTLE. Perhaps the wording projects the statement more into the future than was intended. I think Dr. Peto is referring to trials as done today.

Dr. PETO. That is correct. I believe we must give clinicians a simple message and make it clear that they do not need complicated statistics for the trials that are being done now. The methods I have described will suffice.

Dt. STEWART. As a clinician I think it is essential that the results of studies are presented in a manner which the majority of clinicians will understand. Equally important, however, is that in exploring the data the statisticians should use whatever methods available, no matter how complicated.

Dr. TSIATIS. Please do not misinterpret my point. I like simple methods too but I cannot support a statement suggesting statistical knowledge has reached its limits.

Dr. LAGAKOS. Certainly one must present results in an understandable fashion. There has been a communication problem in the past and both sides must overcome this. However, I feel that the statistical challenges for analysing clinical trial data are considerable. New methods must be developed that apply more to individual patients (e.g. the analysis of precursors of failure, tumour markers etc.), the questions must be simple but not always 'Is A better than B?'

Dr. HAYBITTLE. Yes, there are two levels of analysis. The first is in comparing treatments directly where the simple logrank type methods will suffice. The second is in sorting out prognostic factors where it might be better to use a Cox model analysis.

Dr. PRENTICE. I would like to agree with Dr. Stewart and also to agree with the point that Dr. Haybittle just made.

In regard to Dr. Stewart's comment I think it is essential, in a discussion of statistical methodology, to distinguish between methods that are useful for the exploration and analysis of the data and methods that are useful for the presentation of the findings. In particular, the latter may need to be restricted and tailored to the

background of the target audience.

Noticeably absent from Dr. Peto's position statement on useful statistical methodology for the analysis of nonsequential clinical trials are methods based on the Cox regression model, which is sometimes referred to as the proportional hazard model. I would like to list some reasons for the inclusion of this as well as other regression techniques.

One of the major reasons is that just raised by Dr. Haybittle concerning the 'sorting out' of prognostic factors. It also relates to Dr. Baum's comment that the purpose of a clinical trial is to learn more about the disease in addition to the investigation of treatment strategy. Regression techniques provide a systematic approach to the identification of important prognostic factors. While it may, in principle, be possible to arrive at similar conclusions by considering each potential prognostic factor singly along with stratification on other factors, such a procedure is likely to be so slow and awkward as to be inadequately carried out in most circumstances. Even for the basic comparison of treatments, a regression analysis would be useful for the selection of prognostic factors which may then be used to form strata for the purpose of further analysis and data presentation. In situations with many potential prognostic factors and relatively few data points, such as occur in our work with bone marrow transplantation for leukemia, it is a virtual necessity to use regression techniques for data analysis and also, sometimes, for data presentation. I have found the clinical and basic science people I work with to be generally interested in and receptive to the use of such techniques once the power and versatility of the methodology becomes apparent.

The use of regression techniques permits one to go beyond a simple significance test to compare treatment groups, to the production of a statement that attempts to quantify a treatment difference. For example, a Cox-type analysis may give rise to a statement of the form 'the risk for local recurrence in treatment B is estimated to be 1.7 times that in treatment A'. The method generalizes naturally to allow such a relative risk to vary with the length of time under study.

A regression analysis can also give one a formal basis for comparing treatment effects among subsets of the patient population. Additional components can be readily incorporated in the regression vector in order to test for interactions between auxillary variables and treatment effects while adjusting for other prognostic factors. One must,

however, be cognizant of multiple significance testing problems if interactions are sought in an exploratory manner. Let me conclude by referring back to a statement that Dr. Peto made yesterday indicating that logrank and Cox procedures are identical. I would describe this relationship differently; namely, that any analysis possible with the logrank procedure is a special case of an analysis eminating from the proportional hazard model. The basic logrank test, the logrank test stratified on prognostic factors, the logrank test based on time period subdivision and other time-dependent versions of the logrank test (as discussed by Dr. Crowley) are all score tests for $\underline{\beta} = \underline{0}$ that arise from the proportional hazards technique. As indicated above, however, the proportional hazards technique allows one to go much further.

Dr. HAYBITTLE. Dr. Peto, would you accept any of the comments concerning the value of the Cox model?

Dr. PETO. Of course I accept the Cox model is of value, although I do believe you can often get nearly as much from a simpler analysis. My aim is to try to send a clear message from the statisticians to the clinicians. This is that there are only a few statistical techniques which they need to understand well for the purpose of interpreting clinical trial data and there is not a large amount of complicated mathematics required.

Dr. PRENTICE. I agree, complex methods are usually not necessary in reporting results. However I must say on the few occasions I have felt I had to use a regression analysis for reporting I have found a good understanding among clinicians.

Dr. PETO. Let me summarise my feelings then: the Cox model is a desirable technique but not essential in sorting out treatments in clinical trials.

Dr. STEWART. Surely that is not all we want to obtain from clinical trials. It is a basic feature of trials that we can find out more about the prognostic variables in relation to the treatments, not just a bonus.

Dr. LAGAKOS. One of the biggest mistakes in analysing clinical trial data statistically is to stop the analysis too soon. We are obliged to look at all aspects and find any suggestion of interactions

with whatever method is appropriate. If necessary these can be reported separately from the main points of the study. Too many trials are summarised as 'not significant' or '.05 significant' and I think this is a problem.

Dr. NISSEN-MEYER. I must put a specific question to Dr. Peto. Could you please explain the advantage of the logrank method over the old method of using lifetables and Greenwood's formula (Cutler and Ederer 1958)?

Dr. PETO. The anvantage is usually quite definite although there are a few special occasions with the 'old method' works best. This old method compares the proportion of survivors at just one fixed point in time. It gives you no idea of the nature of the curves elsewhere or of whether the point you have chosen gives an atypical result. Clearly this is important as the data are likely to be sparse if, as is often the case, the region of interest is somewhere towards the end of the curve. However, if you only want to compare the proportions of survivors at a single time then the old method is still best.

The logrank and Cox's method compare the overall shapes of the curves, therefore testing the differences through the whole time period. If there is really a difference in efficacy between the two treatments being compared you are more likely to detect it if you use logrank methods to compare your actuarial survival curves.

Dr. NISSEN-MEYER. Then a difference found by the old method will also be shown by the logrank test.

Dr. PETO. You would almost certainly find that both statistical methods yielded a similar degree of significance in the particular case of your cyclophosphimide data, so for your data the answers would be the same.

Dr. PRENTICE. Dr. Peto's explanation was very clear. However, there are other methods of accumulating differences. If you suspect a treatment difference that is most evident in early survival, tests such as the rank sum or Wilcoxon may be more efficient.

Dr. PETO. This could also be done by dividing the time period into two regions, one early and one late, and using the logrank test on each region separately.

Dr. v. FOURNIER. Do I understand that we can use age corrected survival rates like Duncan and Kerr?

Dr. PETO. You must calculate your survival curves by the actuarial method. Duncan and Kerr did not. The details of survival curve calculations and most of the methods I have described are in our British Journal of Cancer paper (Peto et al. 1976, 1977).

Dr. v. FOURNIER. I have learnt much from the statisticians here but there still remains the time problem. Can statistics help us to reduce the timescales needed for a trial, perhaps to five years rather than ten years?

Dr. HAYBITTLE. No, you can always put forward a model which says if the curve does one thing for five years it does something different for the next five years. Only by extrapolation from a similar group of patients could you infer anything.

Dr. PRENTICE. You can only compare ten year survival rates if you have a ten year follow-up.

Dr. RIBEIRO. This is especially so in breast cancer. In other diseases like lung cancer perhaps you could say that in two years the recurrence rate was so low that meaningful conclusions could be made.

Dr. HAYBITTLE. We have now reached the last question, 'How can we construct an efficient experimental design for a controlled clinical trial on the treatment of breast cancer?' The definition of the hypotheses and the subject matter have both been discussed. An obvious problem remaining is how to find a reasonable number of patients in a relatively short term. Trials are thus likely to be multicentre and we earlier agreed that in that case central randomisation is needed. Dr. Baum, you advocated regional trials, could you summarise your reasons again?

Dr. BAUM. There are two, firstly for easier communications and secondly to establish a group identity. Unfortunately, finding enough people with similar ideas in one area is difficult and I have never been successful in forming a regional group.

Dr. HaYBITTLE. Dr. Peto could you comment on the Medical Research Council's (MRC) trials?

Dr. PETO. The most successful MRC trials have been ones in childhood leukemia and most childhood leukemia patients in Britain are in trials. This is partly due to the enthusiasm and good salesmanship of the organizer, Dr. Kay. The trials are organized partially on a regional basis in that there is a representative in each one of the fourteen National Health Service regions in Britain. All activities are organized via this representative and to a certain extent there are regional loyalties.

Dr. RIBEIRO. Leukemia is a special case because it needs a special expertise to treat it. People go to the groups with a good reputation and these are in the MRC trials. Almost every surgeon thinks he is an expert in how to treat breast cancer and this is why they do not go in for trials.

Dr. PETO. If you can get large enough numbers into trials regionally then a regional trial is a good idea. Recently, however, there was a survey of clinical trials in Britain and most regional trials were found to be of a grossly inadequate size. This was true of some national trials but it was striking that there were only one or two regional trials which were of a reasonable size.

Dr. STEWART. I agree with Dr. Peto but again breast cancer may be a special case. The West Midlands Breast Cancer Trial is one example with adequate numbers. We hope to organize one in Scotland but it is difficult to standardise the treatments in a way which will attract a large number of surgeons. You must trade off having a compact retional study with having to widen the entry requirements.

Dr. HAYBITTLE. Could anybody comment on trials in the USA?

Dr. LAGAKOS. Yes, I would like to make a very practical point. If anybody here is interested in starting a multicentre trial and has no experience it would be advisable to first visit an experienced centre such as ours. You would quickly see where the sources of possible problems are and be able to avoid them.

Dr. ROBERTS. I wonder how the statisticians feel about the quality of the data they get. Is there any attempt to make checks, to exert some form of quality control on the data we present you?

Dr. RIBEIRO. This is one of the advantages of having a regional trial like at the Christie Hospital in Manchester. The records are readily available and the surgeons taking part are known to you personally. It is quick and easy to check on a problem. Also the type of operation used can more easily be standardised and tailored to the surgery policy in the particular region.

Dr. PRENTICE. Much of the data we coordinate is of a slightly different nature from the clinical trial data discussed here but certainly we are very concerned about its quality. Although the data items involved are very basic we have found a great deal of variability in the quality of the abstraction and coding of certein data items. Our attempts at quality control have included, for example, the issuance of the same dummy patient records to each data abstractor in our system in order to assess the reliability of the data which is returned to us.

Dr. LAGAKOS. A big difference between trials in the USA and here in Europe is the amount of money spent. Usually each institution involved in our trials has a paid data manager whose job it is to control the quality of data. At the statistical centre further checks are also made. The data managers have workshops where they meet and discuss any problems and in general improve communications. I feel the benefits of employing such people are well worth the extra expense. You can ensure consistent and quality data by hiring one person in each centre to fill out the forms.

Dr. PETO. In this context I must say that I believe we collect far too much data in trials at present. It is obvious to the people completing the forms that the majority of data will never be used. Consider the Milan study; what records did they really have to complete in order to produce the published data? Surely just the kind of treatment, whether they have alopecia, conjunctivitis etc, number of nodes at presentation, premenopausal or postmenopausal, whether they are dead and whether they have recurred at follow-up. This is all the statistician would need recorded.

Dr. ROBERTS. I think we collect too much data but it is more complicated than Dr. Peto suggests. Consider as an example the recording of conjunctivitis. It is not a clearly defined variable and two clinicians would give you different figures for a set of patients. We

should first be worried about the inconsistencies in our definitions of such variables.

Dr. LAGAKOS. We cope with disputes such as whether there is a response or not by having a study chairman who can look at all the case records. I agree that we tend to include some questions simply because we feel we may need the answer at some future time. Often one does not have enough time to do all the analyses one wants and perhaps not all of the data are used. My feeling is that you should record as little as possible, but endpoints such as tumour recurrence are quite complex. Records must be thorough enough to allow a proper review and to establish that similar measures are being used in several different institutions.

Dr. RIBEIRO. I think you must be careful that you do not lose valuable information.

Dr. ROBERTS. I do not want to record more data, I just feel that we should be more accurate in what we record by making sure the questions are well defined at the beginning of a study.

Dr. RIBEIRO. I agree, we must start right away or the situation will become worse.

Dr. HAYBITTLE. Thank you. I believe we have now covered all the important points in the initial questions. Therefore I would like to formally bring the discussion to a close and thank all participants. I should also like on behalf of the participants to thank our hosts Prof.Immich and his staff, for organizing what I think we all agree has been a most interesting and worthwhile symposium.

Appendix

This comprises the questions which were sent out with the original invitation to participants upon which the discussion was based.

(1) What shall we measure on whom? Why? (M.Schneiderman)

(2) How does the clinician define the effect of a treatment? How can you put such a definition in a mathematical model?

(3) Are the results of statistical tests convincing in such a way that the clinician is obliged to apply a particular treatment only?

(4) How should we handle departures from the process of randomisation (treatment withdrawals, changes of treatment) in a mathematical model?

(5) In which situations should we use
   (i) statistically routine methods (e.g. logrank test, rank sum tests)
   (ii) the regression model of Cox,
   (iii) competing-risk models?

(6) How should we construct an efficient experimental design for a controlled clinical trial on treatment of breast cancer?

CONTROLLED ILLNESS PROCESSES WITH INCOMPLETE INFORMATION.

A RE-EXAMINATION OF THE BASIC HYPOTHESES IN CONSTRUCTING STOCHASTIC MODELS FOR CLINICAL RESEARCH

Petre Tautu

Institute for Documentation,Information and Statistics
German Cancer Research Center
Heidelberg

> It is now,as it was then,as it may ever be, conceptions from the past blind us to facts which almost slap us in the face.
>
> W.A.Halsted

1.Introduction

If we concentrate our attention on the title and the purposes of this symposium,we have to remark that,in fact,we are faced with two main questions :

(1)How does the researcher define and evaluate the effect(s) of a treatment for a chronic disease (e.g.cancer),

(2)What must be the contribution of mathematics to the scientific investigation in the clinical area such that some practical consequences may result.

Let us start with the second question. By 'mathematics' we mean here 'mathematical model',that is,generally speaking,a simplification and a specification in mathematical terms of some aspects of an observed empirical process. It has already been said (Kline,1953) that,primarily,mathematics is a method of inquiry known as postulational thinking : "The method consists in carefully formulating definitions of the concepts to be discussed and in explicitly stating the assumptions that shall be the basis for reasoning. From these definitions and assumptions,

conclusions are deduced by the application of the most rigorous logic man is capable of using." Given this art of drawing conclusions and suggestions for new experiments,"modern science triumphs by virtue of mathematics". It is desirable that modern clinical research appropriately uses this rich and fruitful method of inquiry even if the price to be paid might be the modification of some authoritative ideas or methods of observations and measurements. Construction of appropriate mathematical models describing the natural history of a chronic disease together with the consequent medical decisions and actions would be one of the most recommendable methods of inquiry and explanation in clinical studies.

Of course,this basically implies a great deal of work in formulating precise definitions and concepts. Words such as "favourable state", "active disease","death by cancer" - or other vague clinical expressions- must be replaced by exact formulations. This also intimates the effort of changing some of our thoughts about the occurrence and evolution of a chronic disease. It seems that some views in this domain may be attributed to the old mental scheme about acute diseases with few states and rapidly oriented transitions. Erroneous extrapolations and quasi-exclusive lighting of some facets in the theory of chronic diseases are indeed the effects of this kind of reductionism. The search for one causal agent in cancer or atherosclerosis,the exclusive importance conferred to environmental factors,the use of elementary dose-response relationships for multifactorial pathologic phenomena are some examples of an inappropriate inquiry in the field of chronic diseases. The existence of a misleading 'lecture' of pathologic phenomena justifies the motto above.

For instance,"the major act of faith in the epidemiology of the last thirty years has been that methods which have worked admirably in infectious disease and industrial intoxication can be successfully applied to chronic diseases - cancer,arteriosclerosis,and the like" (Murphy,1978). Even if we can fit some observations with such methods, it must be recalled that "excellence of fit of some function does not guarantee insight". The likeness between 'acute' and 'chronic' diseases is also justified in a recent paper (Barrett-Connor,1979) but not all arguments can be accepted. For instance,it is said that "heart disease, usually classified as chronic,is acute to those myocardial infarct victims who die before reaching the hospital". The confusion between a 'chronic' disease and the sudden (='acute') occurrence of a terminal event cannot help us to acquire a deep insight into these two classes of pathologic processes. It is also said that the word 'chronic' means

slow progression and long duration,but "progression" and "duration" are not defined. As a matter of fact,both are the conventional features of a complicated biological process : the transitions through different states of a chronic disease are made after a long but variable sojourn time and the next state (step) is,in some sense,conjuncturally picked out. We presume that a chronic pathologic process progresses through more optional states than an acute process : a chronic disease with a single preclinical state does not actually exist and this is clear when we look,for instance, at the initial "slow progression" of a cancer disease. The long sojourn time in clinically unobservable (occult) states and the time taken to choose the next state are perhaps important distinctions between 'acute' and 'chronic' diseases. It is known (see e.g.Baessler,this volume) that the time interval between the previous diagnosis of the carcinoma in situ and the subsequent diagnosis of an invasive carcinoma ranges from a few to 20 or 30 years.

If we represent the course of a chronic disease by a directed graph we are often driven to distinguish different principal paths and to interpret them as possibly different clinical entities (see Note I). What is called "tuberous sclerosis" may comprise many distinct diseases (Murphy,1978) and the idea that the natural history of breast cancer represents a single phenomenon is undeniable an anachronism (Fisher and Gebhardt,1978). William of Ockham is best known for his maxim : "It is vain to do with more what can be done with fewer" - which actually is a fruitful principle when it is not trivially used. The necessity of a clear insight into the complexity of biological processes urges us to do with more what we cannot do successfully with fewer.

2.The design of the present paper

We start with two working hypotheses :

(H1)The statistical analysis of complex (biological) processes must finally be the statistical inference for the corresponding stochastic models.

In his 1959 paper on the impact of the theory of stochastic processes on statistics,M.S.Bartlett stated that the "correct specification" of statistical problems "has only become possible in terms of stochastic processes". He pointed out that a feature of probability theory to be noted is "its even wider unifying character for the applications...stimulating their theoretical as well as more empirical statistical aspects". The construction of stochastic models for chronic diseases was from the beginning paired with their statistical analysis.

In 1950,J.Neyman devoted 27 pages of his textbook on statistics and probability theory to the evaluation of 'competing risks' inferred from a simple stochastic model for illness and death. A year later,in his joint paper with E.Fix (see Section 2),he also deduced maximum likelihood estimates of the transition probabilities and proposed the average length of 'normal' life as a measure of the success of a treatment. Some recent contributions are the evidence of a new stage in this field of research.

(H2) Strictly speaking,'breast cancer' does not exist. In point of fact,'breast cancer' is a group of malignant diseases of different cellular origin and history having breast(s) as the organ of localization.

The breast is a complex structure comprising a wide variety of cells of ectodermal and mesodermal origin. The histological appearances of tumors reflect both the kind of cells from which they originate and the special relationship which exists between the epithelial and connective tissue elements (Roe,1979). It has been shown (Adair et al.,1975) that some - unfortunately rare - breast carcinomas (comedocarcinomas, papillary,medullary carcinomas,etc.) have the best survival rate while lobular or high-grade ductal carcinomas show the lowest rates. Thus, 'breat cancer' is only "an eponym label to designate a heterogeneous group of cancers of the breast residing in a heterogeneous group of women...Not only is breast cancer a pathologic heterogeneous disease, but it evokes heterogeneous tumor and host-immune responses as well" (Fisher and Gebhardt,1978).

Our fundamental position is that a correct and useful mathematical model for a chronic disease must necessarily contain more information than that is directly accessible to clinical observations. As P. Carbone said, "good clinical research is good tumor biology"(1977).

In this sense we agree with the idea that we have to construct 'deep' mathematical models (Blumenson and Bross,1969) which can explain the dynamics of a disease as well as the response to our (rational) therapeutic actions and which can also guide us for further research. It has been already remarked (Bross,1972) that the natural history of breast cancers has many unexpected twists and turns but the 'surface' (=clinical) events in this disease show few simple patterns. A deep model must be supported by the information obtained at the next underlying level of that of interest. For example,if "tumor present" is a 'surface' variable,a 'deep' one must contain at least histomorphological and/or biochemical knowledge.

The patient's state is in a mathematical model an event which

needs a correct description with a substantial observational language. The description of a clinical state as e.g. "alive,under treatment" adopts the most superficial observational terms : it leads to the construction of an elementary,'discursive' model,without any cognitive function,by a forced homogenization of patients. Similarly,the state "active tuberculosis" is a markedly heterogeneous classification "since it includes both the patient who is nearly dead of tuberculosis and the patient who is nearly well" (Alling,1958). J.Berkson and R.P.Gage (1952) have rightly pointed out that in some cases the true cause of death is by no means so easy to determine as the fact of death.

It is the hope that the knowledge of tumor growth or cell kinetics parameters will be useful in constructing 'deep' models for clinical evolution or the effect of radio- and chemotherapy. Thus,in the Blumenson-Bross model for breast cancer (see Paragraph 5.3) the most important parameter is the 'tumor doubling time' which actually is a 'net' doubling time for it takes into account host defenses,cell deaths,and the possibility that only a fraction of cells are dividing. We should interpret this approach in a wide sense as follows : The clinical process as considered by L.E.Blumenson and I.D.J.Bross is actually a stochastic process driven by another stochastic process,namely a tumor growth process. The features and the intensity of the latter should explain the evolution of a cancer disease as well as the therapeutic results. I.D.J.Bross suggested (1972) the construction of interconnected mathematical models in order to bridge "the present clinic-laboratory gap in biomedical research".

From a mathematical viewpoint,compounding stochastic processes is not an easy task but such processes are very important and strongly needed for some biological applications (Tautu,1976). In some particular cases notable results are obtained (doubly stochastic Poisson processes, random hazard doubly stochastic Poisson processes,compund Poisson processes,etc.). The simplest possibility is to replace in some situations the considered underlying (basic) random variable by its expected value (see e.g.models A and B in the epidemiologic work of Lewis,1975). But if the Ockham's razor is too frequently used,there is a great risk of changing the deep model back to a surface one. If in the case of the Blumenson-Bross model one avoids the assumption about the existence of cell loss,cell migration or resting fraction,the tumor growth model is an exponential one. We ought to recall here the arguments of W.Feller (1966,p.52) against the universal but misbehaved use of the 'law of logistic growth' : "Theories of this nature are short-lived because they open no new ways,and new confirmations of the same old thing soon grow boring. But the naive reasoning as such has not been superseded by common sense,and so it may be useful to have an explicit demonstration

of how misleading a mere goodness of fit can be."

From a biological point of view,the misappropriate use of a sharp Ockham's razor leads to frustrating results,if any. For instance,there is no clinical schedule using combinations of drugs which has taken advantages of the differences in cell kinetics of tumor and normal tissue (see e.g. van Putten,1974;Tannock,1978).

It has been said that a deep model is much more vulnerable to empirical refutation than a corresponding surface model (Blumenson and Bross,1969). Indeed,the boundaries of the parameter space are specified in a relatively simple way for a superficial model,while for the deep model it is subject to various restrictions : internal (interrelationships of the parameters required for the model to be consistent) and external (imposed by the scientific context).

In what follows we are going to re-examine some mathematical models for cancer diseases,especially for breast cancers and to suggest a new approach,namely the construction of controlled (semi-Markov) illness processes (with partial information). The problem of comparison of treatments and of survival is accordingly deduced. Some mathematical details and comments are presented in Appendices and Notes. It is the hope that in some of them there is substance for new research.

A part of such investigation has been reported in previous papers (Pesky and Tautu,1970;Iosifescu and Tautu,1973;Tautu,1973,1977, 1978).

3.Preliminaries : the Fix-Neyman model for chronic diseases

3.1. In 1949,J.Neyman participated in the Annual Meeting of the American Statistical Association,particularly in the Biometric Section, where a session dealt with the problem of long time follow-up in morbidity studies (see Biometrics 1950,5,345). The session speakers (P.Densen, H.Dorn,T.Harris et al.) published their papers in "Human Biology" (see References). J.Neyman recorded that his interest was later reinforced by acquaintance with statistical studies of the effects of the treatment of 'breast cancer',reported by two different research groups. He was thus brought in a natural way to a Bernoulli-like idea,namely that "in evaluating the effectiveness of a method of treating cancer it is natural to consider the net risk of death from cancer with the risk of being lost or of dying from other causes eliminated".

J.Neyman and E.Fix constructed a stochastic model in 1951 with four possible states for a cancerous patient : 'normal' life, alive but treated for cancer, state of being lost after recovery (either through

difficulties in tracing a patient,or death from other causes than cancer), and death from cancer. In Fig.1 we introduce a graphical representation of this model. Each circle represents one state: state O is of 'normal' life (also including patients "cured from cancer" whether the recuperation is real or only clinically apparent), state 1 is the state being under cancer treatment. There are two 'absorbing' states : state 2 for lost patients or death from other causes and state 3 of death for cancer ('operative' death). The possible transitions between these states are indicated by arrows ; under the hypothesis that alive treated patients may suddenly die from other causes,an arrow must be also drawn from state 1 to state 2.

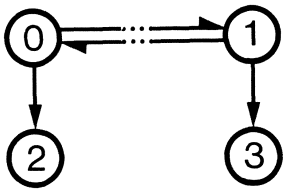


Figure 1

Transition diagram for the Fix-Neyman model

O:normal life ; 1:alive under treatment ;
2:lost after recovery or death from other causes ; 3:death from cancer

The main critique of this model concerns the undifferentiated state 1 and its connection (signifying recovery) to state O. Not only do the patients begin the treatment in different stages of their cancer disease but also the response to the same treatment may be different. Actually individuals in state 1 represent a nonhomogeneous group of patients whose history and objective state was not taken into consideration. The existence of concealed intermediary states is indicated in Fig.1 by the interrupted arrows. Moreover,if one wants to delineate the results of cancer therapy,state 3 is also undifferentiated. It might be a set of states as e.g. dead,no known recurrence ; dead,established recurrence ; dead,uncertain recurrence,etc. In Appendix 1 mathematical arguments for the construction of a state space with a sufficient number of states are given.

3.2. Let us define the Fix-Neyman model as a continuous-time Markov chain $\{\xi(t)\}_{t\geq 0}$ with discrete state space $S=\{0,1,2,3\}$. If at a given time s the patient is in a transient state $i\in S_1=\{0,1\}$,then at a

later time t it may be either in the transient state $j \epsilon S_1$ or in a final (absorbing) state $a \epsilon S_2 = \{2,3\}$, independently of the patient's history until time s. Let us introduce the conditional probabilities :

$$p_{ij}(s,t) = P\{\text{patient in state } j \text{ at time } t \mid \text{patient in state } i \text{ at time } s\}\ ,\ i,j \epsilon S_1\ ,\ s<t,$$

and

$$p_{ia}(s,t) = P\{\text{patient dead in death state } a \text{ at time } t \mid \text{patient alive in state } i \text{ at time } s\}\ ,\ i \epsilon S_1\ ,\ a \epsilon S_2\ ,\ S = S_1 \cup S_2.$$

The p_{ij}'s are called 'illness transition probabilities' and the p_{ia}'s 'death transition probabilities' (see Chiang,1968).

We assume in what follows that the Markov process is homogeneous, i.e.the transition probabilities $p_{ij}(s,t)$ depend only on the difference t-s,so that $p_{ij}(s,t) = p_{ij}(0,t-s) = p_{ij}(t-s)$. The p_{ij}'s satisfy the fundamental conditions,namely $p_{ij}(t) \geq 0$, $\sum_j p_{ij}(t) = 1$, and $\sum_k p_{ik}(s) p_{kj}(t) = p_{ij}(s+t)$.

We introduce now the following postulates :

(i) If at epoch t the process is in state $i \epsilon S_1$,the probability that between t and a small time interval $t+\Delta t$ it goes in state $j \epsilon S_1$ equals $\lambda_{ij}\Delta t + o(\Delta t)$.

(ii) The probability of a transition from $i \epsilon S_1$ to a death state $a \epsilon S_2$ in the same short time interval $(t,t+\Delta t)$ is $\mu_{ia}\Delta t + o(\Delta t)$.

(iii) The probability that during $(t,t+\Delta t)$ more than one change occurs is $o(\Delta t)$,where $o(\Delta t)$ approaches zero at a higher order than Δt.

In the simple case studied by E.Fix and J.Neyman,the expressions for the illness transition probabilities are obtained as

$$p_{ij}(t) = \sum_{\substack{u,v=1 \\ u \neq v}}^{2} \frac{\lambda_{ij}}{\rho_u - \rho_v} \exp\{\rho_u t\},\ i \neq j \epsilon S_1 \qquad (1)$$

The death transition probabilities p_{ia} are

$$p_{ia}(t) = \sum_{\substack{u,v=1 \\ u \neq v}}^{2} \frac{(\rho_u - \lambda_{jj})\mu_{ia} + \lambda_{ij}\mu_{ja}}{\rho_u(\rho_u - \rho_v)} \exp\{\rho_u t\} - 1\ ,\ i \neq j \epsilon S_1\ ,\ a \epsilon S_2 \qquad (2)$$

The ρ's are the two real roots of the characteristic equation

$$\begin{vmatrix} \rho - \lambda_{ii} & -\lambda_{ji} \\ -\lambda_{ij} & \rho - \lambda_{jj} \end{vmatrix} = \rho^2 - (\lambda_{ii} + \lambda_{jj})\rho + (\lambda_{ii}\lambda_{jj} - \lambda_{ij}\lambda_{ji}) = 0.$$

The reader can find the details in any classical book on stochastic processes (e.g. Chung,1967,p.134;Cox and Miller,1965,p.178).

3.3. E.Fix and J.Neyman called both transition probabilities (1) and (2) the 'crude' probabilities in order to point out that the transitions from state 0 to state 1 or vice versa are affected by some competing risks. For instance,the transition probability p_{10} (recovery probability) is influenced by the two death eventualities : the transitions from state 1 to state 3 or to state 2,that is,death by cancer during the treatment or death by other causes. The 'net' probability of recovery can be obtained if we assume that the intensities μ_{12} and μ_{13} are zero. The 'net' risk of death from cancer can be deduced from (2) by putting $\mu_{02}=$ $=\mu_{12}=0$.

We must firstly notice that a 'competing risk' is some state (or a set of states) contrasting to a certain representation of linearity. If from an arbitrary state 1 (its real signification is now without interest) the process can go to,say,states 2,3 or 4,the competing risks are states 3 and 4,assuming that the transition to be taken into account is 1→2. Let us consider the hypothetical Markov chain (Fig.2) constructed by B.Berlin,H.Preisler and E.Sid (Neyman,1975). When the elementary path of interest is 0→1→2→5, the states 4,3 and 6 are the competing risks - or,the deviations from the path points (0,1,2,5).

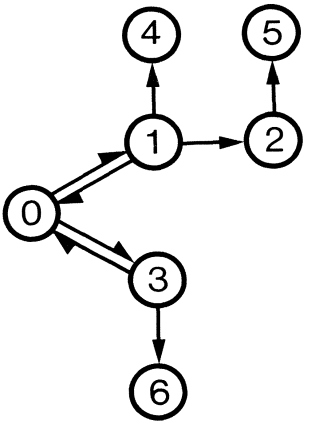


Figure 2

Transition diagram for an illness - death model

0:healthy state ; 1:benign tumor ; 2:malignant tumor ; 3:other diseases (lethal); 4:death from benign tumor; 5:death from malignant tumor; 6: death from other diseases

Introducing certain arbitrarily chosen values of transition probabilities, the authors estimated (by using the nonparametric Kaplan-Maier 1958 method) the 'net' probabilities for death by other diseases,death from benign tumors and death from cancer. The estimations for the first two probabilities do not fit the true curves. The conclusion is that "a correct technique based on an unrealistic model of a phenomenon cannot be expected to yield realistic results"(Neyman,1975).

The problem of competing risks can be generally formulated in probability terms as follows. Suppose that states 3,6 and 4 (Fig.2) form a prohibited set H of states. For a finite Markov chain $\{\xi_n\}_{n\geq 0}$ with stationary transition probabilities p_{ij},the probability

$$p^{(n)}_{ij,[H]}=P\{\xi_n=j,\ \xi_\nu\notin H,\ 0<\nu<n\,|\,\xi_o=i\} \qquad (3)$$

is the conditional probability that starting from initial state i the process reaches state j in n steps without entering in the set H of prohibited states during the steps 1,2,...,n-1. Probabilities (3) are called 'taboo' probabilities and can be interpreted as transition probabilities in a modified chain in which the prohibited set H has been made a closed set (see Chung,1967,p.45;Cox and Miller,1965,p.107; Gihman and Skorohod,1975,I,p.111). A 'net' probability is,in fact,a kind of 'taboo' probability. (See Appendix 2 for some interesting developments.)

3.4. The concepts of 'crude' ('dependent':Zwinggi,1945;'influenced':Hoem,1969) and 'net' ('independent':Zwinggi,1945;'absolute':Jordan, 1952;'partial':Hoem,1969) probabilities are basic for the stochastic approach to the Daniel Bernoulli 1760 problem : If in a given population a disease could be eliminated,what would be the effect,in probability terms,onthe population mortality structure at different ages ? This problem has been of great interest and importance to actuaries for over 100 years. It is,therefore,not surprising that some of their results were rediscovered by the probabilists (compare e.g.some results obtained by J.Neyman (1950) with the Du Pasquier's formula (1913) or C.L.Chiang's 1961 proportionality assumption with J.Meier (1940) or T. Greville (1948) approach). The 'disability theory' (Du Pasquier,1913) may be viewed as the primordial scheme of an illness-death model.

In the long run the present problem of competing risks traces over the old model for acute infectious diseases. In the Bernoulli's and Laplace's model there is an immediate heavy mortality affecting those individuals who catch smallpox and also a specific action (vaccination) which reduces this mortality. To all appearances there is no

analogy between the competing risks model for smallpox and a competing risks model for a group of chronic diseases. The first argument is that the model may be deserving if there exists a strong difference in the contribution to mortality between a disease of interest D_1 and a set D_2 of certain other diseases. If D_2 has only a mild influence,then the difference between the independent and the influenced probabilities of D_1 can be trivial (Neyman,1972).

The second argument is that the main aim of Bernoulli's approach was to evidence the vaccine effectiveness by estimating the 'partial crude' probabilities of death when smallpox is eliminated. We still not possess for chronic diseases specific therapy or preventive actions which can eradicate them. In that case we cannot substantiate and use the old metaparadigm "one factor - one disease" but the hypothesis "multiple factors - a family of diseases". Given the network of "causation factors",the elimination of a disease can alter the contribution of other diseases. D'Alembert wrote (1761) at some length on the difficulty of comparing "un danger présent"(=death from vaccination) with "un danger éloigné"(=death from smallpox within a given period). This may be our case when the present danger is,say,breast cancer and the remote danger is irradiation (mammography). The occurrence of a second malignancy (e.g.acute leukemia) by X-rays treated patients with breast cancers is a rare event and apparently "un danger éloigné".

The third argument is the confusion made between the probability of dying from a disease and the event itself. In the 18th century the identification could be possible in susceptible host populations,for a fatal disease like smallpox. At the present day there is little clinical evidence to support the selection of a single underlying cause of death (see e.g.Tolley et al.,1978). Moreover,it was already emphasized that probability of dying and death rate are two different from one another; the first cannot even be unambigously calculated from the latter (Keyfitz and Frauenthal,1975).

In fact,mortality is a 'surface' variable which gives us no direct information about the complicated pattern and interrelationships between diseases. The validity of the hypothesis of risks independence (or of additivity of their 'forces of mortality') depends "in part on the disease in question and its complex relationship with other diseases in the particular host population"(Chiang,1961). A 'deep' model for competing risks must be a patient-oriented stochastic illness model (Tautu, 1978) which may be,for instance,a particular type of chain. (See Appendix 3). A.Tsiatis also suggested (1978) the abandon of the old approach "in favor of a stochastic model that describes the underlying mechanism of interaction of different diseases".

A fresh interpretation of the Fix-Neyman model as a competing risks model may be found in M.Gail (1975).

3.5. At the time J.Neyman began his study on the medical follow-up problem,a generally accepted measure of therapy effectiveness has been the probability that a treated patient survives five or more years with cancer. It is easy to discern that in the Fix-Neyman 1951 model the appropriate measure is the average sojourn time in state O : a treatment is more beneficial than another if the patient remains longer time in the 'normal life' state (see again Fig.1). Apparently,A.Berger and R.Z. Gold (1961) were the first who tried to estimate the survival times for the Markovian model built up by E.Fix and J.Neyman. The reader will find in Note II a general treatment for the 'expected duration of stay' in a certain state of a Markovian illness process.

It is now obvious to analyze survival by introducing the event invalidating it,namely the 'failure'. The occurrence of a failure cuts off the trajectory of a given process. If this process represents e.g. the life of an individual,his death from some cause is a failure. Some specifications are often required. D'Alembert has already distinguished the 'physical' life from the 'real' life : A woman without breast cancer has a real life in respect to this disease ; a woman alive with breast cancer stage IV has a physical life. This,if the process we consider represents the real life of a woman,the initiation of a breast tumor is interpreted as a failure. One must realize that the word 'failure' may designate any other "soft end-point which may be not immediately obvious to the patient or study subject"(Johnson and Koch,1978),e.g.the occurrence of a tail when the coin tossing process describes the head succession, the fail in an examination,bankruptcy,etc. In a patient-oriented illness model the occurrence of a pathologic event is a failure. If the process of interest is a 'clinical process' (Iosifescu and Tautu,1973,p.240),i.e. a stochastic process starting with the clinical diagnosis of a (chronic) disease and ending with the patient's death,this terminal point is a failure. Then the survival time is the time interval between diagnosis and death,that is,the lifetime of the random clinical process. Our efforts are to make it as long as possible. This can be gained over (1)by displacing it proximally by introducing specific and effective screening procedures,and (2)by displacing it proximally by using an efficient therapy.

These remarks are of importance,I suppose,for the accurate interpretation of the survival time. We said that it is the lifetime of a Markov clinical process $\{\xi(t)\}_{t\geq 0}$,that is,a random variable defined as

$$\zeta(\omega) = \inf\{t : \xi(t,\omega)=a\}, \ \omega\in\Omega$$

with Ω the space of elementary events and $a\notin S$ an absorbing state. $\zeta(\omega)$ can be interpreted as the exit time from the state space S. Since $\{\zeta(\omega)>t\} = \{\xi(t)\neq a\}\in\mathcal{B}_t$,it follows that $\zeta(\omega)$ is a random time with respect to the σ-algebras $\{\mathcal{B}_t\}_{t\geq 0}$ (see e.g.Gihman and Skorohod,1975,II, p.89). This formal definition has to point out that the survival time as the lifetime of a stochastic clinical process is not a self-explanatory empirical concept but a general probability notion in relation to a system of sets of clinical events.

3.6. It is well known that the probabilistic failure behavior of a process is generally specified in terms of a function f(t) which is a probability density function defined on the real axis and representing the instantaneous failure at time t :

$$f(t) = \lim_{\Delta t\downarrow 0+} \frac{P\{t<X\leq t+\Delta t\}}{\Delta t} \ ,$$

where X is a nonnegative random variable,the failure time. In the medical terminology f(t) is the 'death intensity function'. It has the following properties : $f(t)\geq 0$ and $\int_0^\infty f(t)dt=1$. Corresponding to f(t) is its distribution function F(t),usually called 'cumulative distribution function',which gives the probability that a failure occurs by time t. Then the complementary 'survivor' function is $\bar{F}(t)=P\{X\geq t\}=1-F(t)=$ $=\int_t^\infty f(x)dx$. The Cox's proportional hazard model can be specified in terms of $\bar{F}$ as

$$\bar{F}(t|\underline{z}) = \exp\{ -\Sigma(t) \exp(\beta\underline{z})\} \ , \qquad (4)$$

where $\exp\{-\Sigma(t)\}=P\{X\geq t|\underline{z}=\underline{0}\}$ is a base line survivor function (Kalbfleisch, 1978). In Appendix 4 the reader will find some details about the random process $\Sigma(t)$ as the subordinator of the process $\bar{F}(t)$.

3.7. We close this section with the following recent theorem given by C.L.Chiang (1979).

Theorem. If an individual is subject to a chronic disease able to be represented as a chain of s-1 illness states and one absorbing (death) state accessible from any illness state,then his survival time ζ has the density function

$$f_\zeta(t) = \exp\{\lambda_{11}\int_0^t \phi(x)dx\}\mu_1\phi(t) + \sum_{j=2}^{s-1}\left[\lambda_{12}\cdots\lambda_{j-1,j}\mu_j\phi(t)S_{ij}\times\right.$$

$$\left.\times \exp\{\lambda_{ii}\int_0^t \phi(x)dx\}\right] \ ,$$

where $\lambda_{ij}\phi(t)$ and $\mu_i\phi(t)$, $1\leq i,j\leq s-1$, are the transition rates with

$$\lambda_{ii}\phi(t)=-\left[\lambda_{i,i+1}\phi(t)+\mu_i\phi(t)\right] \ , \ 1\leq i\leq s-1,$$

and $\lambda_{ss}\phi(t)=-\mu_s\phi(t)$. The function $\phi(t)$ is a function of time t at which the transitions take place, with assumption $\int_0^\infty \phi(t)dt=\infty$, and

$$S_{ij} = \sum_{i=1}^{j}\left[\prod_{\substack{k=1\\k\neq i}}^{j}(\lambda_{ii}-\lambda_{kk})\right]^{-1}.$$

Corresponding to $f_\zeta(t)$ there are its distribution function

$$F_\zeta(t) = \frac{\mu_1}{\lambda_{11}}(\exp\{\lambda_{11}\int_0^t\phi(x)dx\}-1) + \sum_{j=2}^{s-1}\left[\lambda_{11}\cdots\lambda_{j-1,j}\mu_j S'_{ij} \times \right.$$

$$\left. \times(\exp\{\lambda_{ii}\int_0^t\phi(x)dx\}-1)\right]$$

with $F_\zeta(\infty)=1$, and the expectation

$$E[\zeta] = -\frac{\mu_1}{\lambda_{11}}\int_0^\infty\exp\{\lambda_{11}\int_0^t\phi(x)dx\}dt - \sum_{j=2}^{s-1}\left[\lambda_{12}\cdots\lambda_{j-1,j}\mu_j S'_{ij} \times\right.$$

$$\left.\times\int_0^\infty\exp\{\lambda_{ii}\int_0^t\phi(x)dx\}dt\right] ,$$

where

$$S'_{ij} = \sum_{i=1}^{j}\left[\prod_{\substack{k=1\\k\neq i}}^{j}(\lambda_{ii}-\lambda_{kk})\lambda_{ii}\right]^{-1}.$$

4.Some advances in the mathematical theory of chronic diseases

4.1. The Fix-Neyman model can also be treated as a compartment model (e.g.Matis and Wehrly,1979) especially when we are going to investigate the "population process". A state is defined to be a compartment which contains a (random) number of patients who move from compartment to compartment and from some compartments out of the system (death). We have then to consider a stochastic multi-compartment system (see e.g. Thakur et al.,1973;Faddy,1976,1977).

Let us consider,for example,a 5-compartment system as follows: compartment 0 contains individuals in apparently health state and having no risk for cancer ; compartment 1 contains apparently healthy individuals but with high risks for cancer : they will move to compartment 2 when they will be tumor hosts. All individuals being in compartments 0, 1 and 2 can die from other causes than cancer (compartment 3) ; individuals in compartment 2 can also die from cancer (compartment 4). Thus the input in compartment 3 is composed with individuals moving from all compartments,excepting compartment 4 which has as input only a fraction of individuals being in compartment 2. This model (Tolley et al.,1978) intends to represent the evolution of gastric cancer in a closed population and its goal is "to synthesize a number of findings into a comprehen-

sive model of carcinogenesis in general human populations" and at the same time to determine the combination of the factors which may "realistically characterize the latent stages of carcinogenesis".

The following postulates are introduced :

(i) If a tumor begins to grow in the age interval $(a,a+\Delta a)$, given that the individual was in compartment 1 at age a, the transition intensity to compartment 2 is $\lambda_1(a)\Delta a+o(\Delta a)$.

(ii) The probability that a patient dies from cancer in age interval $(a_o+t, a_o+t+\Delta t)$, given that tumor started its growth at age a_o and the patient was alive in compartment 2 at age a_o+t, is $\lambda_2(t)\Delta t+o(\Delta t)$.

(iii) The probability that an individual alive at age a will die from other causes than cancer in the age interval $(a, a+\Delta a)$ is $\mu(a)\Delta a+o(\Delta a)$.

(iv) The probability that during the interval $(a+t, a+t+\Delta t)$ more than one change occurs is $o(\Delta t)$.

H.D.Tolley and his colleagues introduced in their model some hypotheses on carcinogenesis and tumor growth, namely the multiple hit hypothesis and the assumption that tumor grows following an exponential law. Thus, the intensity λ_1 has the following explicit form :

$$\lambda_1(a) = k_1 a^{m-1},$$

where k_1 is a constant subsumming the probabilities of each of m mutations and m-1 is the 'characteristic power' of the tumor of interest. If m mutations are required for tumor onset, then the age-specific incidence rate of tumors is proportional to the (m-1)th power of age (see e.g.Doll,1971). It was suggested (Stocks,1953) that for a gastric tumor seven mutations could be necessary. Thus, λ_1 is the incidence rate of gastric cancer.

For the death intensity λ_2 the following relation is proposed :

$$\lambda_2(t) = k_2 \exp\{\beta t\},$$

where t is the time interval since tumor starts, k_2 is a constant relating tumor size to the risk of death by cancer and β is a parameter governing tumor doubling time (under the hypothesis that the tumor grows exponentially). Further, the competing risks effects are given by

$$\mu(a^*) = k_3 \exp\{\gamma a^*\},$$

where a^* equals a or a_o+t (depending on what compartment, 1 or 2, individuals came), k_3 is a constant relating the effects of 'biological wear' to μ, and γ is a parameter designed to represent the rate of accrual bio-

logical wear. More precisely,γ is a parameter of the Gompertz function. Clearly,$\mu(a^*)$ is the actuarial 'force of mortality' from all other causes than cancer.

The hypotheses used for λ_1 and λ_2 can be easily invalidated in spite of their workability (see e.g.Paragraph 5.4).

4.2. An alternative description of this model by using the sojourn times in different compartments is also given (Tolley et al.,1978). Let us consider the following sojourn (waiting) times :

W_1,the age at which the malignant tumor initiates,i.e.the sojourn time in compartment 1 ;

W_2,the age with cancer,i.e.the sojourn time in compartment 2,until the entry in death compartments 3 or 4 ;

W_3,the age of death due to tumor load,i.e.$W_3=W_1+W_2$;

W_4,the age of death due to other causes.

It must be pointed out that W_3 and W_4 are not exactly sojourn times but 'holding' times because the next states (compartments 3 or 4) are selected. Therefore,the mathematical definitions are not identical with those for W_1 and W_2. The alternative description strongly suggests the construction of a semi-Markov compartment model (see e.g.Marcus and Becker,1977).

4.3. A non-Markovian model called 'fatal illness' and analogous to the Fix-Neyman model has been reported by D.R.Cox and H.D.Miller in their book (1965,p.253). The authors consider a 3-state model,without introducing the state of death from other causes and,thus,the problem of competing risks is neglected. The basic assumption is that there exist a holding time X in the healthy state O until the clinical detection of a tumor (state 1) and a failure time Y which is the time spent in state 1 until death (state 2).

The probability of having a tumor at time $(t+\Delta t)$ given that the patient had this tumor at time t equals $1-h(y)\Delta t+o(\Delta t)$,under the condition that the patient has been in state 1 (cancerous) for time Y=y. Hence the required probability is obtained by integrating the hazard function h(y) over the distribution of the time that the patient has been spent in state 1 up to time t. To obtain this latter distribution is "as complicated as solving the process as a whole" (Cox and Miller, 1965).

In order to get solutions for such a non-Markovian process,it must be transformed in a Markovian one. For example,instead of considering cancer disease as one single state,one assumes that the disease has k stages (taken in series or in parallel),the duration of each stage being independently exponentially distributed. The non-Markovian problem is thus transformed to a Markovian problem. Another modality is the

the addition of a sufficient number of supplementary variables to define the states of a process. We obtain a multidimensional Markov process.

As a result of the non-Markovian approach we deduce that (i)a Markovian illness model must have a sufficient number of states,(ii)these states must be well-defined,on occasion by the aid of some supplementary (explanatory) variables. The vector $\underline{z}$ of concomitant variables in (4) is an example. Recently,it was suggested the introduction of multiple time scales which may play the role of time-dependent covariates (Farewell and Cox,1979). For instance,the failure (incidence) time for breast cancer may be recorded as chronological age but also as age since some major hormonal event occurs,e.g.the birth of the first child.

5.On some cancer disease models

5.1. By studying the survival curve for cancer patients following tratment,J.Berkson and R.P.Gage (1952) constructed a simple stochastic model,starting with the common assumption that patients with a specified cancer are,before treatment,all subject to the effect of two forces of mortality : by other diseases and by cancer. These forces act independently and simultaneously. The model consists of two transient states (1,living with uncured cancer; 2,living with cured cancer) and two absorbing states (3,death by cancer; 4,death by ordinary causes). Since there is no communication between states 1 and 2,the process is finally the sum of two subprocesses,one with state space {1,3,4} and other with {2,3,4}. It reminds the structure of the compartment model in Paragraph 4.1,with the distinction that the latter was oriented to the tumorigenesis process while the first intends to show the effect of cancer therapy.

But the authors emphasized,with full knowledge of the case,that their assumptions oversimplify the situation :"the presence of cancer influences the probability of death from other causes...there are specific seasonal characteristics of the different causes of death so that they do not operate with strict simultaneity...the effect on mortality is more complicated than the sharp dichotomization pictured..."

5.2. In 1968,B.A.Barron and R.M.Richart constructed in an inherent way a Markov model for the natural history of cervical carcinoma. The state space of this model contains four states : 0,normal; 1,dysplasia; 2,carcinoma in situ; 3,invasive cancer. This is the widely accepted paradigm of the disease (see Dunn,1953;Younge,1965). Although the model has been considered valid,other observations (Coppleson and Brown,1975) show that it is "totally incompatible" with some observed data,essentially because the Barron-Richart model is a time-homogeneous Markov

chain. For instance,the transition probability p_{01} (normal→dysplasia) rises sharply with age and then falls to a low value. Conversely,the reverse transition probability p_{10} does not change greatly with age. Before age 43,the values of p_{12} rise and then fall but after this age they increase sharply with age. Also,the values of p_{23} increase monotonely with age. As L.W.Coppleson and B.Brown remarked,the transition probabilities p_{12} are high in the early years,fall appreciably in the forties,and then rise again. This would suggest the existence of two cancerogenetic processes,"one occurring clearly during the reproductive period,the other occurring after the menopause". Moreover,it is suggested that carcinoma in situ should be "a mixture of two different lesions", a benign condition that spontaneously regresses and occurs mainly before age 50 and a premalignant condition seen mainly after age 50 when it does not regress and does jump to state 3,invasive carcinoma.

All these are arguments demanding consideration for the construction of nonhomogeneous illness processes (Appendix 5).

5.3. In 1969,L.E.Blumenson and I.D.J.Bross published a model for breast cancer rationally based on a simple TNM-like classification. This is apparently the first stochastic model which takes into account the cancer disease as a whole and not only the tumor growth : the presence of node metastases is an element of the disease state as well as the possible recurrence after treatment (surgery). Then a state can be represented by a vector of three integer variables indicating tumor size, number of metastatic nodes and occurrence of recurrence. For instance, the vector (0,0,0) represents the state of a patient apparently free of cancer : either non-cancerous or cancerous after a radical ('proper') operation before recurrence occurs. The vector (1,0,0) indicates the disease state with a small breast tumor and the vector (1,1,0) indicates the presence of a small tumor with 1-3 'positive' nodes,etc. In Fig.3 we introduce a graphical representation of a segment of the natural history by using the above staging. The dotted arrows show the dynamics when cancer therapy (e.g.surgery) is applied.

The Blumenson-Bross model is not easy to understand owing the complexity of postulates,the fuzzy notation and the nature of model chosen. The authors consider three time intervals : T_1,the interval between the initiation of carcinogenesis to the tumor detection, T_2,the interval between detection and removal of a tumor (the 'delay' time), and T_3,the interval between the initiation of carcinogenesis and the initiation of the remote metastasis. If t is the time the tumor is detected, the probability that it is discovered prior to T_2 can be written as

$$P\{t<T_2\} = 1 - \exp\{-\frac{1}{2} k\beta T_2^2\} ,$$

where β is the tumor doubling time (a constant under the hypothesis of exponential growth) and k a proportionality factor.

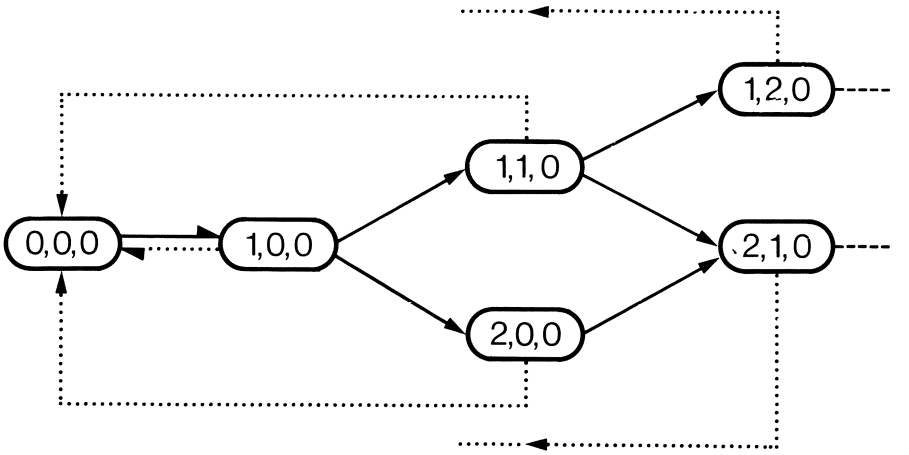


Figure 3
Transition diagram for the Blumenson-Bross model

The first vector component indicates the tumor size (0:no tumor; 1:small tumor,<5 cm diam.;2:large tumor,>5 cm diam.);the second component indicates the metastatic lymph nodes (0:no cancerous nodes;1:one to three positive nodes; 2:more than three positive nodes);the third component indicates the recurrence (0:not clinically detectable;1:clinically detectable).

Also,the conditional probability that at time t there are no positive nodes if there is a delay of T_2 may be written as

$$P\{\xi(t)=(\cdot,0,\cdot)\,|\,t\geq T_2\} = \exp\{-\alpha(T_1+T_2)\},$$

where α is a measure of the susceptibility of the nodes to tumor involvement and ξ(t) is the state of the considered pathological process at time t. The probability that the patient is in state (1,0,0) at time t is

$$P\{\xi(t)=(1,0,0)\}=\left[P\{\text{patient enters the study}\}\right]^{-1}\int_0^7 P\{\xi(t)=(\cdot,0,\cdot)\,|\,t\geq T_2\}P\{\xi(t)=(\cdot,\cdot,0)\,|\,t\geq T_2\}dP\{t\geq T_2\}.$$

The integration domain [0,7] is expressed in terms of doubling times. It is approximated that 30 doublings are required before the primary tumor reaches a detectable size. Seven other doublings are necessary for the transformation of a small tumor into a large one. For this reason the length of the time intervals are measured on a 'tumor scale' in units of

tumor doublings.

The consideration of this clinical parameter leads the authors to the idea that there exist at least two different breast cancer diseases, one with 'fast' doubling time and another with 'slow' doubling time. But the range of this variable is between 1.2 days and 900 days (see Table 3 in Gullino,1977) or,estimated from mastectomy scar recurrences, between 25 and 240 days (Pearlman,1976;Ackerman and Katzenstein,1977). It is rationally impossible to separate 'slow' and 'fast'breast cancers on such a wide-range basis.

5.4. A digression is here inevitable. It is well-known that the assumption of a constant tumor growth rate - following the exponential law - is not correct. Tumor cells do not grow synchronously and only a portion of the cell population is dividing. Moreover,10^9 malignant cells are not necessary to form a 1-cm breast tumor because mammary carcinomas are,in fact,constituted by only 21-65% of cells,the remainder being stroma (Underwood,1972).

In a recent paper,R.J.Gratton et al.(1978) used seven functions with three parameters which all fitted well tumor growth data. Their conclusion is that a great many curves may be use to give apparently reasonable representations of the data "but when the equations are used to estimate the specific growth rate of the tumor the results may be misleading and the standard error found by the curve fitting method may lead to a considerable underestimate of the possible error". Hence it appears that there are strong biological and mathematical arguments against the exponential growth hypothesis and further against the extrapolation procedure for evaluating the birth of the first cancer cell (A stochastic approach is suggested in Note III.)

We must mention for the young reader the discussion of W.A.O'N Waugh to a paper delivered by M.Zelen in 1966. He has been struck by the results obtained when an exponential function is used. He pointed out that in certain conditions a stochastic growth model - which is more realistic from the biological point of view - can have a sort of insensitivity to some changes in certain probability distributions underlying it. If W is (as usual in the theory of branching processes) a random variable which can be thought of as accounting for the early stochastic fluctuations of a cell population,its distribution is distinctly insensitive to the functional form of the generation time distribution if the variance of the cell lifelength is fairly small (Burnett-Hall and Waugh, 1967).

The main problem is,therefore,what kind of data can a mathematical model take into its quantitative assumptions and further what kind of

data must validate it. Being a theoretical construction,a mathematical model must abstract its concepts from different empirical sources - e.g. clinical but morphological,pathological but physiological,etc. Recently, W.F.Eddy (1979) claimed that "the only purpose of models is to make formal implications. For far too long,statisticians have concentrated on fitting models to data. And,for reasons I don't fully understand,they have 'tested' the parameters of these models. The relevance of models come only from their implications and the interpretation thereof."

5.5. A recent model of G.M.Tallis et al.(1979) represents the evolution of breast cancer as a function of primary tumor size and the degree of local nodal involvement. The flow diagram for their Model 2 is given in Fig.4.

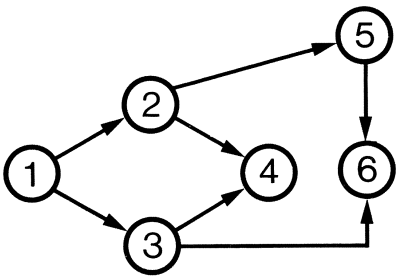


Figure 4

Flow diagram for the breast cancer Model 2 (Tallis et al.,1979)

1:Primary tumor only; 2:Primary tumor + (1-3 nodes); 3:Primary tumor + (distant metastases); 4:Primary tumor + (1-3 nodes)+(distant metastases); 5:Primary tumor + (4 or more nodes); 6:Primary tumor + (4 or more nodes) + (distant metastases)

Let $\lambda_{ij}(v), 1 \leq i,j \leq 6$, be the rate of formation of metastases for a primary tumor of volume v. The case $\lambda_{ij}(v) \equiv \lambda_{ij}$ corresponds to the assumption that the metastatic activity is proportional to the rate of growth of tumor volume. In order that the rate λ of metastatic activity is kept constant,one assumes that there is a transformation of v, $T[v(t)]$ for which the usual Markov requirements are satisfied. Because in practice one measures the tumor diameter,let $v^{1/3}=x$. Then,the probability that the process is in state 1 (primary tumor,no metastases) is

$$p_1(T) = e^{-(\lambda_{12}+\lambda_{13})T} .$$

Looking at Fig.4 we see that the probability of being in state 2 is

$$p_2(T) = \frac{\lambda_{12}}{(\lambda_{24}+\lambda_{25})-(\lambda_{12}+\lambda_{13})} \left[\exp\{-(\lambda_{12}+\lambda_{13})T\} - \exp\{-(\lambda_{24}+\lambda_{25})T\}\right].$$

The situation when the primary and secondary metastases grow independently is also studied (see Karlin and Taylor,1975).

5.6. An attempt to use an amended TNM classification is briefly presented in our book (Iosifescu and Tautu,1973,p.243). When cancer is staunchly considered to be a disease we should not focus on tumor growth only but also on other symptoms of gravity. We mentioned in 1973 the hormonal status,ignoring at that time the early reports on oestrogen receptors (e.g.Korenman and Dukes,1970). Such a factor of gravity (Knight et al.,1978) as well as the indication of a constellation of biochemical markers (Coombes,1978) must be introduced in the definition of clinical states,particularly when the effects of therapeutic actions are to be investigated. Prolonged and unnecessary chemotherapy is harmful for patients and could be avoided if no fall in marker level was observed. Furthermore,drug dose could be adjusted according to marker response since this may occur before the disease is visibly altered in extent.

We must also distinguish between invasive and noninvasive primary breast tumors (see the Postsurgical Treatment Pathologic Classification of the 1977 TNM System given by the American Joint Committee for Cancer Staging and End Results Reporting (AJC) together with UICC). Small invasive cancers (<0.5 cm diam.) are associated with axillary metastases in 15% of cases while noninvasive ductal carcinomas have nodal metastases only in 5% of cases. The immunologists suggested a new designation for a nodal status N,namely N^+,indicating a progressive weakening of the local immune response : It is hypothesized that immune deficiency could be ascribed to a 'paraneoplastic' syndrome,that is,a biologic syndrome directly or indirectly controlled by the tumor itself (Israel,1978).

Thus,improving the original idea,each state of disease is to be specified by a complex vector of integer variables : for instance, (100001...) will indicate the presence of a single and small breast tumor,without detectable positive nodes,without detectable distant metastases,in a woman without pregnacy or breast feeding,without pregnancy-associated macroglobulin (PAM),but with dimethylguanosine in urine,etc.

5.7. The first semi-Markov model for a malignant disease has been constructed by G.H.Weiss and M.Zelen (1963,1965) for treated lymphocytic leukemia. The considered process has six states : 1,initial relapse state (condition of patient on entering study); 2,first partial remission

(also including subsequent relapse); 3,second partial remission (including subsequent relapse); 4,first complete remission (including subsequent relapse); 5,second complete remission (including subsequent relapse); O,death by cancer. We present in Fig.5 our simplified version which clearly shows the cycling between remissions and relapses. The final state 5 receives its input only from the relapse state 4. In order to show that the model takes exclusively into consideration the clinical evolution under therapy,a state O (apparent health) is connected by a dotted arrow to the initial relapse state.

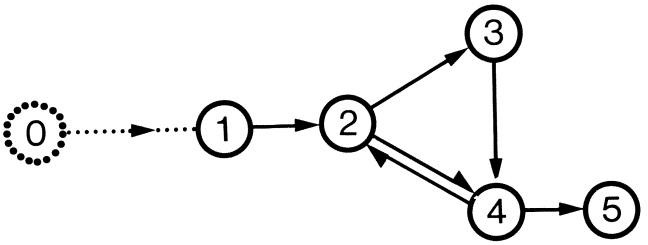


Figure 5

The minimal state space for the Weiss-Zelen semi-Markov model

O:health; 1:First relapse; 2:Partial remission; 3:Complete remission; 4:Relapse; 5:Death

It is perhaps useful to mention that the semi-Markov processes were introduced by P.Lévy (1954) and W.L.Smith (1955) simultaneously at the International Congress of Mathematicians held in Amsterdam. At the same time,L.Takâcs (1954) introduced essentially the same type of stochastic process,applying it to some problems in counter theory. The general theory of semi-Markov processes has been developed by R.Pyke (1961 a,b).

The process rules given by W.L.Smith are as follows. Consider a Markov process $\{\xi(t)\}_{t\geq 0}$ with state space $S=\{1,\ldots,s\}$. Now,

(i) If a transition has just occurred in which $\xi(t)$ enters state $i\in S$ at time t,the probability is p_{ij} that the next transition is into state $j\in S$. Successive selections of states constitute independent random trials.

(ii) If a transition into state i occurred at t=x,and it is given that the next transition will be into state j,$i\neq j$,then that transition will occur at time $x+\tau_{ij}$,where τ_{ij} is a random variable,the "wait in i conditional upon j",and $P\{\tau_{ij}\leq t\}=F_{ij}(t)$,with $F_{ij}(\infty)\leq 1$.

(iii)The successive waits involved,as the process develops,are independent random variables ; none of them is zero with probability one.

(iv)The process $\{\xi(t)\}_{t\geq 0}$ is taken as continuous to the right.

The random variable τ_{ij} plays naturally an important role in the Weiss-Zelen model. If,for instance,i=3 and j=4,then τ_{34} can be a measure of therapy effectiveness : it is the holding time in a state of complete remission before a relapse occurs. Another measure may be the time to reach state 4 for the first time.

Looking at the data obtained from the records of 54 patients (Frei et al.,1961),we notice that the sojourn time in state 1 is between 2 to 12 weeks while the sojourn time in the second relapse state is between 10 and 24 weeks. The patients with lymphocytic leukemia were initially treated with methotrexate and in a second phase with 6-mercaptopurine.

As for the other illness models presented here,the main critique one should bring to the Weiss-Zelen model is the insufficient number of states in S and their vague definition - which ignores the characteristics of chronic lymphocytic leukemias. Therapy should have a goal of achieving a complete remission (state 3 in Fig.5) in order to (i)eliminate a leukemic clone of cells and allow the habitual production of normal lymphocytes,(ii)re-establish immune function and reduce infectious morbidity,(iii)re-establish normal marrow function,and (iv)permit consequently longer survival. There is no state defined in view of these pathologic events. We also know that chronic lymphocytic leukemia is a disease of variable course and survival. In some studies the median survival time of patients with 'active' disease is calculated and is equal to 27 months,while for patients with 'indolent' disease it is 52 months. A clinical staging classification in relationship with survival might be necessary (see Rai et al.,1975).

5.7. A 3-state semi-Markov model has been presented by S.W.Lagakos in 1976. The states are linearly connected : 1,alive,without progressive disease; 2,alive but having previously experienced progressive disease; 3,dead. The model was used for analyzing survival data but without making real profit of the theory of semi-Markov processes. A generalization of an s-state process is presented in (Lagakos et al.,1978).

Note.We must mention that one of the first semi-Markov models in medicine has been constructed by E.B.Perrin and M.C.Sheps (1964). It dealt with the process of female fertility. In his 1971 book,R.A.Howard proposed a problem by considering the dynamics of the patients in a hos-

pital as a 4-state semi-Markov process (p.684). For coronary patients that dynamics has been studied by E.Kao (1972,1974). A renewal model for chronic diseases has been reported by S.M.Berman (1965) who analyzed the age of onset and death at a given point in time. Information on the survival times is obtained from age-specific incidence and prevalence rates. (The calculation of the relative risk of a disease from prospective and retrospective studies,by using a Fix-Neyman model,has been suggested by H.Sugiyama (1961).)

6.Controlled stochastic illness processes with incomplete information

6.1. We initially assume that the most appropriate stochastic model able to describe the evolution of a chronic disease is a semi-Markov model. Then we say that our disease-oriented model is the SM-illness process. We keep the term 'illness' up instead of 'disease' though there is a tendency to consider illness as a human event and disease as a biologic one,i.e.illness consists of an array of discomfort and psychosocial dislocations resulting from interaction of a person with the environment (Barondess,1979). We use the two terms interchangeably.

Let us firstly admit that the disease states are carefully specified (possibly as vectors like in Paragraph 5.6) and form a finite state space S. It must contain the helathy state O,some preclinical states and clinical states,that is,the natural history of the considered disease. We suppose that a patient jumps from one state of this disease to another state with real-valued,random sojourn times in between. Let us denote by ξ_n the state space entered after the n-th sojourn and by τ_n the random duration of this n-th sojourn. The successive states are to form a time-homogeneous Markov chain $\{\xi_n\}_{n\geq 0}$ and the sequence $\{\tau_n\}_{n\geq 1}$ of sojourn times for a real valued process defined on the same probability space and linked with the Markov chain as follows :

$$P\{\xi_n=j\ ,\ \tau_n\leq x \mid \tau_1,\tau_2,\ldots,\tau_{n-1},\xi_o,\xi_1,\ldots,\xi_{n-1}=i\} =$$

$$P\{\xi_n=j\ ,\ \tau_n\leq x \mid \xi_{n-1}=i\} = p_{ij}F_i(x) = Q_{ij}(x),$$

where F_i is a proper right-continuous distribution function for each i. The matrix $\underline{Q}(x)=(Q_{ij}(x))$ is called a semi-Markov matrix. The random variables $\{\tau_n\}$ are then conditionally independent given the values of the chain. In fact,

$$P\{\tau_m\leq x_m,\ 1\leq m\leq n \mid \xi_o,\xi_1,\ldots,\xi_{m-1}\}=\prod_{m=1}^{n}P\{\tau_m\leq x_m \mid \xi_{m-1}\}.$$

The sequence $\{\tau_n\}$ is called a sequence of random variables defined on a Markov chain (Fabens and Neuts,1970;Resnick and Neuts,1970;Wolfson,1977) or a chain-dependent process (O'Brien,1974;Denzel and O'Brien,1975). A process $\{\xi_n,\tau_n\}$ of the above type is called a (J,X)-process (Pyke,1961; Hatori et al.,1967;Janssen,1969;Oprisan,1976). This process satisfies

$$\tau_o=0 \quad \text{a.s.}$$

$$P\{\xi_o=i\}=p_i \quad , \sum p_i=1.$$

Put

$$\{\xi(t)\}=\xi_k \quad \text{if} \quad \sum_{m=0}^{k-1}\tau_m < t \leq \sum_{m=0}^{k}\tau_m \ .$$

The process $\{\xi(t)\}_{t\geq 0}$ jumps at the time $\sum_{m=0}^{k}\tau_m$,k=0,1,2,... and is constant between jumps. $\xi(t)$ is a semi-Markov process. From its definition it is clear that $\xi(t)$ is a process which,on the one hand,possesses the properties of a Markov process and,on the other hand,possesses the properties of renewal processes,because the epochs of successive transitions into a fixed state form a renewal process. In my 1977 paper I have defined the SM-illness process by means of the triple $(S,\underline{p},\underline{F})$,where S is the finite state space,$\underline{p}$ is the initial probability vector and $\underline{F}$ is the matrix of the cumulative probability distributions of holding times τ_{ij}, $i,j\in S$. The reader will find in Appendix 6 a more formal definition which leads to important probabilistic and statistical developments.

6.2. Cancer is,as R.Bellman said (1973),"an ideal field for control theory because on the one hand the basic physiological processes are so bound up with these mathematical ideas [of control theory] , and on the other hand because all aspects of prevention,diagnosis,therapy, and hospital operation involve allocation of resources,decision-making and risktaking,which is to say control theory". The construction of a controlled SM-illness process is our main scope.

With this purpose in view,we require

(i)the modification of the state space S by including some new disease states that shall specify the effects of therapy,that is,the medically modified natural history of the disease. We have to insert,for instance,in the definition of a state of a patient under treatment measurable as well as evaluable but nonmeasurable effects of therapy (responses). In many cases the responses are mixed : we record some regressing lesions (e.g.metastases),some progressing lesions,and also some new appearing lesions (Hayward et al.,1978).

(ii)a set A of actions (or 'alternatives') which contain all therapeutic procedures (e.g.cancer surgery,radiation,chemo- and/or

immunotherapy). We then assume that there exists the possibility of operating on the illness process,that is,one assumes that the probability characteristics of the future course of the process can be changed by taking actions of the set A. According to the control problem studied, A may be defined by a list of all potential actions.

(iii)one or more criteria of judging the effects of the decided actions. The problem the physician and the mathematician must to solve together is how to select the actions that will make operation of the illness process "most rewarding". We shall say that the selection of an action when the process enters into a given state is called a 'decision' and the set of decisions for all states constitutes a 'policy' (or strategy,or plan). Then the problem is to find the most rewarding or most profitable policy. It is necessary to specify for each disease and for each situation what means 'most profitable'. It is considered that for finite time intervals,the problem is to find the time-varying strategy that will maximize the 'expected total reward' generated by the controlled process. If this process is an illness process the reward may be, broadly speaking,the price reduction for all medical actions compatible with a certain quality of life. "To begin with,we are matching human life and suffering on one hand and material resources on the other"(Bellman, 1973). Cancer cells killing is not the unique touchstone as long as we are able to understand that we have to do with a disease driven by perturbed control mechanisms that presumably are not the same for every tissue and every cell type. What we really need in cancer therapy is the thorough-going study of cell population biology. All the shortcomings observed in the mathematical models for chemotherapy result "from the paucity of knowledge about many fundamental principles governing the behavior of cell populations"(Donaghey and Drewinko,1975).

The construction of a controlled illness process needs previous clinical experience as well as information about all types of experiments for the conditions of effective therapy,including the 'in numero' experiments of mathematical models. It is known,for example,that if we take into consideration different dose levels and varying time intervals between administration of a combination of four drugs,it is possible to devise a formidable number of protocols (experimental designs) of order of ten millions... For some of these combinations the effects are similar. We refer here only to the paper of B.M.Hancock et al.(1976) where two schedules,FACO (fluorouracil+doxorubicin(adriamycin)+cyclophosphamide+ +oncovin(vincristine)) and FCO gave similar results in advanced 'breast' cancer. The criterion was 50% or greater tumor regression sustained for a month or more. No information is given relating the type of malignant breast disease. Other reports reveal,for instance,that the combination

FCMP (fluorouracil+cyclophosphamide+methotrexate+prednisone) is more effective than FCM (Canellos et al.,1976).

It is perhaps for the first time in the history of therapeutic when a treatment for a chronic disease became impossible without mathematical models and statistical design. Clinical oncology is now an interdisciplinary field.

6.3. There are many mathematical models describing the intracellular biochemical interactions of cancer drugs or their cytokinetic effects. Particularly in the latter the methods of optimal control theory were applied with the view to minimize the size of primary tumor (see e.g. Bahrami and Kim,1975). In spite of their recognized simplicity and theoretical deficiencies,the cytokinetic mathematical models help us, especially when they exploit their predictive function. As T.E.Wheldon (1978) acknowledged,"optimization with more realistic models constitutes a difficult but not intractable problem which,judging by the encouraging results obtained in the present simplistic analysis,deserves further study". It is perhaps trivial to point out that the term 'optimal' has a mathematical meaning and not a clinical one. The deterministic optimal control problem,for example,asks to find the control functions which yield minimal value of an objective criterion (see e.g.Fleming and Rishel, 1975).

It is surprising that the process of resistant cell formation is seldom taken into consideration (Dedrick et al.,1975) as well as the idea induced from some biochemical models (Werkheiser et al.,1973):the quantitative change in parameter value may bring about a qualitative change in the structure of the model. Is it too paradoxically to presume that by chemotherapy in some cases we transform a malignant disease into another (malignant) disease ?

The stochastic aspects of malignant cell growth control have been discussed in (Iosifescu and Tautu,1973,p.253 and 257;Tautu,1978).

6.4. The construction of a controlled SM-illness model moots the optimal problem at the clinical level. When the illness process is in state $j\in S$ and we choose an admissible action $a\in A_j$,three things happen:

(i)the process moves to a new state selected according to the probability distribution $Q_j^a(\cdot)$. If the process remains in j we call that a 'virtual' transition.

(ii)conditional on the event that the new state is $k\in S$,the length of time the process takes to move to state k from state j is a nonnegative random variable with probability distribution $F_{jk}^a(\cdot)$.

(iii)conditional on the event that the new state is $k\in S$,immediately after the transition is completed,we receive a 'reward' whose probabil-

ity distribution is $R^{a}_{jk}(\cdot)$.

A strategy $\underline{\sigma}$ for the control of an illness process is a sequence $\{\sigma_1,\sigma_2,\ldots\}$ of decision rules where the n-th decision rule σ_n tells us how to select an action in A after completion of the (n-1)st transition. More precisely, σ_n is a conditional probability on A given the history H_n of the whole controlled process up and including the epoch of the (n-1)st transition, that is, $H_n=(\xi_1,a_1,t_1,r_1,\ldots,\xi_{n-1},a_{n-1},t_{n-1},r_{n-1},\xi_n)$. That is to assume that given the observed history H_n up to the time of the (n-1)st transition, we choose our n-th action according to the distribution $\sigma_n(\cdot|H_n)$. Thus, on the one hand, the decision σ at each epoch depends on the trajectory of the process uo to this time, and, on the other hand, the probabilistic characteristics of the process depend on the decisions $\sigma_1,\sigma_2,\ldots$ A semi-Markov strategy is a sequence $\{\sigma_1,\sigma_2,\ldots\}$ where each σ_m is a measurable function from S×S into A and $\sigma_m(\xi_1,\xi_m)$ is the action we take at the m-th step if we start in state ξ_1 and the m-th state is ξ_m.

How we rationally select a therapeutic action is an interesting problem which should make the object of a separate study - e.g.Savage's 'sure-thing principle' (Tautu,1973 b).

The problem for the controlled SM-illness processes can be simply stated as follows : find the policy that maximizes the expected total reward. The accurate clinical meaning of 'reward' may be either to gain a certain time or to reach a certain 'favourable' state (e.g.partial or complete remission in leukemias) or both. The dynamic programming solution (with finite horizon) is referred to as 'value iteration' (Howard, 1963;1971). This is approximately the scaffolding of our 1977 paper. Statistically speaking, comparing two controlled illness processes means comparing two strategies for an identical history H.

6.5. We present now more intuitively another possible approach. Let us fix a certain 'favourable' state $\alpha\in S$ and denote $\tilde{S} = S\setminus\{\alpha\}$. We assume now that the transition probabilities are such that $\alpha\in S$ can be reached from every state $j\in S$. The control consists in this case whether a transition shall take place or not. The object is to try to reach state α in the shortes possible time.

Such an approach is very interesting when we look only at the Markov chain associated with the SM-illness process. With each policy used, there is associated a new Markov chain with α as absorbing state. This strategy actually minimizes the hitting time of state α or the first entry in the considered absorbing state. The model was generally constructed for arbitrary Markov chains with countable state space (Kesten and Spitzer,1975).

6.6. The control of an illness process needs exact information about the present state in order to select the most appropriate action. A clinical diagnostic as "no evidence of primary tumor" can hide the existence of a small tumor of 0.5 cm or 1 cm diameter (of weight of 100 μg to 1 mg) which approximately represent 10^5-10^6 cells,most of them being malignant (see the graph 1 in Griswold and Corbett,1976). Similar insufficient information is contained in the diagnostic of "palpably normal axilla" : in about 40% of cases such 'normal' nodes are metastatic.

The clinician is thus in the situation to choose an action under incomplete information,that is,to choose an admissible action for state j while the patient actually is in state k. We have to distinguish between the 'kernel' process ('core' or underlying process)$\{\xi(t)\}_{t\geq 0}$ with measurable state space $(S,\mathcal{S})$ and the observed ('skin') process $\{X(t)\}_{t\geq 0}$ with measurable state space $(S',\mathcal{S}')$. If the kernel process is in state $j\in S$ we can observe a state $i\in S'$ of the skin process.

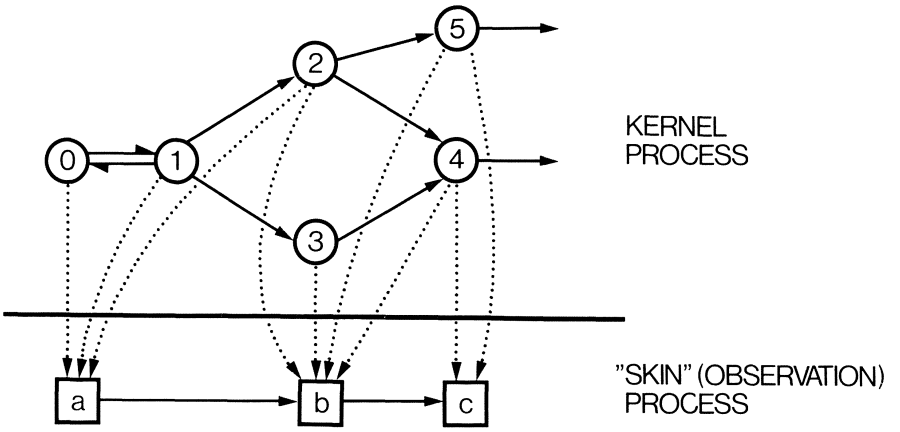


Figure 6

An illustration of state correspondences between kernel and skin processes

It was said above that the selection of an admissible action depends in the semi-Markovian case on the present state and on the number of transitions. This becomes impossible when we have incomplete information about the real SM-illness process,since the present state is incompletely detected and the number of transitions may be unknown. E.B. Dynkin (1966) and Y.Sawaragi and T.Yoshikawa (1970) solved the problem in the Markovian case by means of a set of probability measures that sub-

stitute the unobservable part of the state space. This leads to a modified controlled process with observed history. The main problem is now to show that both models -the model with incomplete information and the modified one- are equivalent with regard to optimality. E.B.Dynkin did not mention if in the skin model there are other policies than in the modified model.

We have,in fact,two alternatives : (i)optimize the skin process under certain conditions concerning the information content of the kernel process (White 1974,1976),or (ii)reconstruct the kernel process from the skin process by introducing some 'near-neighbor rules' (Devore,1973) and re-consider the optimization problem. The second procedure may be attractive (and useful) for a sagacious clinician. For a different but interesting approach the reader is referred to R.A.Howard (1965;1971, p.349 and 829).

R E F E R E N C E S

1.Ackerman,L.V.,Katzenstein,A.L.(1977).The concept of minimal breast cancer and the pathologist's role in the diagnosis of "early carcinoma". Cancer 39,2755-2763

2.Adair,F.,Berg,J.,Joubert,L.(1975).Long term follow-up of breast cancer patients.The 30-year report. Cancer 33,1145-1150

3.Alling,D.W.(1958).The after-history of pulmonary tuberculosis:A stochastic model. Biometrics 14,527-547

4.Bahrami,K.,Kim,M.(1975).Optimal control of multiplicative control systems arising from cancer therapy. IEEE Trans.Autom.Control AC-20, 537-542

5.Barondess,J.A.(1979).Disease and illness-A crucial distinction. Amer. J.Med. 66,375-376

6.Barrett-Connor,E.(1979).Infectious and chronic disease epidemiology: Separate and unequal ? Amer.J.Epidemiol.109,245-249

7.Barron,B.A.,Richart,R.M.(1968,1970).A statistical model of the natural history of cervical carcinoma.I,II. J.Natl.Cancer Inst.41,1343-1353; 45,1025-1030

8.Bartlett,M.S.(1959).The impact of stochastic process theory on statistics. In:Probability and Statistics (U.Grenander ed.),pp.39-49. New York:Wiley

9.Bellman,R.(1973).Cancer,computers,and control theory:A prospectus for the next hundred years. Techn.Rep.RB-73,Univ.of Southern California

10.Berger,A.,Gold,R.Z.(1961).On comparing survival times. Proc.4th Berkeley Symp.Math.Statist.Probability,Vol.IV,pp.67-76. Berkeley-Los Angeles:Univ.California Press

11.Berkson,J.,Gage,R.P.(1950).Calculation of survival rates for cancer. Proc.Mayo Clinic 25,270-286

12.Berkson,R.,Gage,R.P.(1952).Survival curve for cancer patients following tratment. J.Amer.Statist.Assoc.47,501-515

13.Berman,S.(1965).A renewal theoretic model for chronic disease statist-

ics. J.Appl.Probability 2,119-128

14.Black,M.M.,Barclay,T.H.,Hankey,B.F.(1975).Prognosis in breast cancer utilizing histologic characteristics of primary tumor. Cancer 36, 2048-2055

15.Bloom,H.J.G.,Richardson,W.W.,Harries,E.J.(1962).Natural history of untreated breast cancer (1805-1933).Comparison of untreated and treated cases according to histological grade of malignancy. Brit.Med.J. 2,213-221

16.Blumenson,L.E.,Bross,I.D.J.(1969).A mathematical analysis of the growth and spread of breast cancer. Biometrics 25,95-109

17.Bross,I.D.J.(1972).Scientific strategies in human affairs:Use of deep mathematical models. Trans.N.Y.Acad.Sci.34,187-199

18.Bross,I.D.J.,Blumenson,L.E.(1968).Statistical testing of a deep mathematical model for human breast cancer. J.Chron.Dis.21,493-506

19.Burnett-Hall,D.G.,Waugh,W.A.O'N (1967).Sensitivity of a birth process to changes in the generation time distribution. Proc.5th Berkeley Symp.Math.Statist.Probability,Vol.IV,pp.609-623. Berkeley-Los Angeles: Univ.of California Press

20.Canellos,G.P.,Pocock,S.J.,Taylor,S.G.,Sears,M.E.,Klaasen,D.J.,Band, P.R.(1976).Combination chemotherapy for metastatic breast carcinoma. Prospective comparison of multiple drug therapy with L-Phenylalanine Mustard. Cancer 38,1882-1886

21.Carbone,P.P.(1977).Tumor biology and clinical trials. Cancer Res.37, 4239-4245

22.Chiang,C.L.(1961).On the probability of death from specific causes in the presence of competing risks. Proc.4th Berkeley Symp.Math. Statist.Probability,Vol.IV,pp.169-180. Berkeley-Los Angeles:Univ.of California Press

23.Chiang,C.L.(1968).Introduction to Stochastic Processes in Biostatistics. New York:Wiley

24.Chiang,C.L.(1975).Neyman's contribution to the theory of competing risks. A Fix-Neyman model. Proc.8th Intern.Biometric Conf.,pp.133-151. Bucuresti:Ed.Academiei

25.Chiang,C.L.(1979).Survival and stages of disease. Math.Biosci.43, 159-171

26.Chiang,C.L.,Hsu,J.P.(1976).On multiple transition time in a simple illness death process.A Fix-Neyman model. Math.Biosci.30,55-71

27.Chiang,C.L.,Raman,S.(1973).On a solution of Kolmogorov differential equations. Proc.4th Conf.Probability Theory (Brasov,Romania,1971), pp.129-136. Bucuresti:Ed.Academiei

28.Chung,C.L.(1967).Markov Chains with Stationary Transition Probabilities,2nd ed. Berlin-Heidelberg-New York:Springer

29.Coombes,R.C.(1978).Biochemical markers in human breast carcinoma. Invest.Cell Pathol.1,347-362

30.Coppleson,L.W.,Brown,B.(1975).Observations on a model of the biology of carcinoma of the cervix:A poor fit between observation and theory. Amer.J.Obstet.Gynecol.122,127-136

31.Cox,D.R.,Miller,H.D.(1965),The Theory of Stochastic Processes. London: Methuen

32.Çinlar,E.(1969).Markov renewal theory. Advances Appl.Probability 1, 123-187

33.Dedrick,R.L.,Zaharko,D.S.,Bender,R.A.,Bleyer,W.A.,Lutz,R.J.(1975). Pharmacokinetic considerations on resistance to anticancer drugs. Cancer Chemother.Rep.59,795-804

34.Densen,P.M.(1950).Long-time follow-up in morbidity studies:The definition of the group to be followed. Human Biol.22,233-237

35.Denzel,G.E.,O'Brien,G.L.(1975).Limit theorems for extreme values of chain-dependent processes. Ann.Probability 3,773-779

36.Devore,J.L.(1973).Reconstructing a noisy Markov chain using near-neighbor rules. J.Amer.Statist.Assoc.68,599-602

37.Doll,R.(1971).The age distribution of cancer:Implications for models of carcinogenesis. J.Roy.Statist.Soc.Ser.A 134,133-166

38.Donaghey,C.E.,Drewinko,B.(1975).A computer simulation program for the study of cellular growth kinetics and its application to the analysis of human lymphoma cells in vitro. Comp.Biomed.Res.8,118-128

39.Donegan,W.L.,Lewis,J.D.(1978).Clinical diagnosis and staging of breast cancer. Seminars in Oncology 5,373-384

40.Dorn,H.F.(1950).Methods of analysis for follow-up studies. Human Biol. 22,238-248

41.Dunn,J.E.(1953).The relationship between carcinoma in situ and invasive cervical cancer. Cancer 6,873-886

42.Du Pasquier,L.G.(1913).Mathematische Theorie der Invaliditätsversicherung. Mitteil.Verein.Schweizer.Versicherungsmath.8,1-153

43.Dynkin,E.B.(1966).Controlled random sequences. Theor.Probability Appl.10,1-14

44.Eddy,W.F.(1979).Comment to E.Parzen's paper "Nonparametric statistical data modeling". J.Amer.Statist.Assoc.74,124-126

45.Enstrom,J.E.,Austin,D.F.(1977).Interpreting cancer survival rates. Science 195,847-851

46.Fabens,A.J.,Neuts,M.F.(1970).The limiting distribution of the maximum term in a sequence of random variables defined on a Markov chain. J.Appl.Probability 7,754-760

47.Faddy,M.J.(1976).A note on the general time dependent stochastic compartment model. Biometrics 32,443-448

48.Faddy,M.J.(1977).Stochastic compartmental models as approximations to more general stochastic systems with the general stochastic epidemic as an example. Advances Appl.Probability 9,448-461

49.Farewell,V.T.,Cox,D.R.(1979).A note on multiple time scales in life testing. Appl.Statist.28,73-75

50.Feller,W.(1966).An Introduction to Probability Theory and Its Applications,Vol.II (2nd corrected printing). New York:Wiley

51.Fisher,B.,Gebhardt,M.C.(1978).The evolution of breast cancer surgery: Past,present,and future. Seminars in Oncology 5,385-394

52.Fix,E.,Neyman,J.(1951).A simple stochastic model to recovery,relapse, death and loss of patients. Human Biol.23,204-241

53.Fleming,W.H.,Rishel,R.W.(1975).Deterministic and Stochastic Optimal Control. New York-Heidelberg-Berlin:Springer

54.Frei,E.,Freireich,E.J.,Gehan,E.,Pinkel,D.,Hollan,J.F.(et al.)(1961). Studies of sequential and combination antimetabolite therapy in acute leukemia:6-Mercaptopurine and Methotrexate. Blood 18,431-454

55.Gail,M.(1975).A review and critique of some models used in competing risk analysis. Biometrics 31,209-222

56.Gihman,I.I.,Skorohod,A.V.(1975).The Theory of Stochastic Processes. I,II. Berlin-Heidelberg-New York:Springer

57.Gratton,R.J.,Appleton,D.R.,Alwiswasy,M.K.(1978).The measurement of tumour growth rates. In:Biomathematics and Cell Kinetics (A.J.Valleron,P.D.M.Macdonald eds.),pp.325-332. Amsterdam-New York-Oxford: Elsevier-North Holland Biomed.Press

58.Greville,T.N.E.(1948,1949).Mortality tables analyzed by cause of death. Record Amer.Inst.Actuary 37,283-294;38,77-79

59.Griswold,D.P.,Corbett,T.H.(1976).Breast tumor modeling for prognosis and treatment. In:Breast Cancer:A Multidisciplinary Approach (G.St-Arneault,P.Band,L.Israel eds.),pp.42-58. Berlin-Heidelberg-New York: Springer

60.Gullino,P.M.(1977).Natural history of breast cancer.Progression from hyperplasia to neoplasia as predicted by angiogenesis. Cancer 39, 2697-2703

61.Hancock,B.M.,Ribeiro,G.G.,Todd,I.D.H.(1976).Combination chemotherapy of advanced breast cancer-A random trial of two schedules. Clinical Oncology 2,101-104

62.Harris,T.E.,Meier,P.,Tukey,J.(1950).Timing of the distribution of events between observations:a contribution to the theory of follow-up studies. Human Biol.22,249-270

63.Hartveit,F.(1971).Prognostic typing in breast cancer. Brit.Med.J.4, 253-257

64.Hatori,H.,Mori,T.,Oodaira,H.(1967).A renewal theorem for (J-X)-processes. Kōdai Math.Sem.Rep.19,159-164

65.Hayward,J.L.,Meakin,J.W.,Stewart,H.J.(1978).Assessment of response and recurrence in breast cancer. Seminars in Oncology 5,445-449

66.Hoem,J.M.(1969).Purged and partial Markov chains. Skand.Aktuarietidsk. 52,147-155

67.Howard,R.A.(1963).Semi-Markovian decision processes. Bull.Intern. Statist.Inst.40,625-652

68.Howard,R.A.(1965).Dynamic inference. Operations Res.13,712-733

69.Howard,R.A.(1971).Dynamic Probabilistic Systems.I,II. New York:Wiley

70.Iosifescu,M.,Tautu,P.(1973).Stochastic Processes and Applications in Biology and Medicine.II:Models. Berlin-Heidelberg-New York:Springer

71.Israël,L.(1978).An immunologic look at the TNM classification: Therapeutic implications and strategies. Cancer Treatm.Rep.62,1177-1182

72.Israël,L.,Chahinian,P.(1975).Mensuration of response. In:Cancer Therapy:Prognostic Factors and Criteria of Response (M.J.Staquet ed.), pp.37-48. New York:Raven Press

73.Janssen,J.(1969).Les processus (J-X). Cahiers Centre Études Recherche Opér.11,181-214

74.Johnson,W.D.,Koch,G.G.(1978).Linear models analysis of competing risks for grouped survival times. Intern.Statist.Rev.46,21-51

75.Jordan,C.W.(1952).Life Contingencies. Chicago:Soc.Actuaries

76.Kalbfleisch,J.D.(1978).Non-parametric Bayesian analysis of survival time data. J.Roy.Statist.Soc.Ser.B 40,214-221

77.Kao,E.P.C.(1972).A semi-Markov model to predict recovery process of coronary patients. Health Services Res.7,191-208

78.Kao,E.P.C.(1974).Modelling the movement of coronary patients within a hospital by semi-Markov processes. Operations Res.22,683-689

79.Karlin,S.,Taylor,H.M.(1975).A First Course in Stochastic Processes. New York:Academic Press

80.Kesten,H.,Spitzer,F.(1975).Controlled Markov chains. Ann.Probability 3,32-40

81.Keyfitz,N.,Frauenthal,J.(1975).An improved life table method. Biometrics 31,889-899

82.Kline,M.(1953).Mathematics in Western Culture. London:Oxford Univ.Press

83.Knight,W.A.,Livingston,R.B.,Gregory,E.J.(1978).Absent estrogen receptor and decreased survival in human breast cancer. Proc.Amer.Soc. Clinical Oncology 19,392

84.Korenman,S.G.,Dukes,B.A.(1970).Specific estrogen binding by the cytoplasm of human breast carcinoma. J.Clin.Endocrin.Metab.30,639-645

85.Lagakos,S.W.(1976).A stochastic model for censored-survival data in the presence of an auxiliary variable. Biometrics 32,551-559

86.Lagakos,S.W.,Sommer,C.J.,Zelen,M.(1978).Semi-Markov models for partially censored data. Biometrika 65,311-317

87.Lévy,P.(1954).Processus semi-markoviens. Proc.Intern. Congr.Mathematicians 3,416-426

88.Lewis,T.(1975).A model for the parasitic disease bilharziasis. J.Appl. Probability 7,673-704

89.Littell,A.S.(1952).Estimation of the T-year survival rate from follow-up studies over a limited period of time. Human Biol.24,87-116

90.Marcus,A.H.,Becker,A.(1977).Power laws in compartmental analysis.II. Numerical evaluation of semi-Markov models. Math.Biosci.35,27-45

91.Matis,J.H.,Wehrly,T.E.(1979).Stochastic models of compartmental systems. Biometrics 35,199-220

92.Mattson,P.(1977).Some reflections on different disability models. Scand.Actuarial J.,pp.110-118

93.Meier,J.(1940).Zur Theorie der unabhängigen Wahrscheinlichkeiten. Mitteil.Verein.Schweizer.Versicherungsmath.39,53-74

94.Murphy,E.A.(1978).Epidemiological strategies and genetic factors. Intern.J.Epidemiol.7,7-14

95.Neyman,J.(1950).First Course in Probability and Statistics. New York: Holt

96.Neyman,J.(1972).Epilogue of the Health-Pollution Conference. Proc.6th Berkeley Symp.Math.Statist.Probability,Vol.VI,pp.575-587. Berkeley-Los Angeles:Univ.of California Press

97.Neyman,J.(1975).Assessing the chain:Energy,crisis,pollution and health. Intern.Statist.Rev.43,253-267

98.O'Brien,G.L.(1974).Limit theorems for sums of chain-dependent processes. J.Appl.Probability 11,582-587

99.Oprisan,G.(1976).On the J-X processes. Rev.Roumaine Math.Pures Appl. 21,717-724

100.Pearlman,A.W.(1976).Breast cancer-influence of growth rate on prognosis and treacment evaluation. Cancer 38,1826-1833

101.Perrin,E.B.,Sheps,M.C.(1964).Human reproduction:a stochastic process. Biometrics 20,28-45

102.Pesky,M.,Tautu,P.(1970).A Markovian model for the clinical evolution of tuberculosis. 7th Intern.Biometric Conf. (Unpubl.paper,abstract Biometrics 27,257-258)

103.Portugal,M.A.,Falkson,H.C.,Stevens,K.,Falkson,G.(1979).Acute leukemia as a complication of long-term treatment of advanced breast cancer. Cancer Treatm.Rep.63,177-181

104.Pyke,R.(1961 a).Markov renewal processes:Definitions and preliminary properties. Ann.Math.Statist.32,1231-1242

105.Pyke,R.(1961 b).Markov renewal processes with finitely many states. Ann.Math.Statist.32,1243-1259

106.Rai,K.R.,Sawitsky,A.,Cronkite,E.P.,Chanana,A.D.,Levy,R.N.,Pasternack, B.S.(1975).Clinical staging of chronic lymphocytic leukemia. Blood 46,219-234

107.Resnick,S.I.,Neuts,M.F.(1970).Limit laws for maxima of a sequence of random variables defined on a Markov chain. Advances Appl.Probability 2,323-343

108.Rilke,F.,Andreola,S.,Carbone,A.,Clemente,C.,Pilotti,S.(1978).The importance of pathology in prognosis and management of breast cancer. Seminars in Oncology 5,360-372

109.Roe,F.J.C.(1979).Aetiology of breast cancer:A brief review. Invest. Cell Pathol.2,45-53

110.Sawaragi,Y.,Yoshikawa,T.(1970).Discrete-time Markovian decision processes with incomplete state observation. Ann.Math.Statist.41, 78-86

111.Smith,W.L.(1955).Regenerative stochastic processes. Proc.Roy.Soc. Ser.A 232,6-31.(Remarks on the paper "Regenerative stochastic processes" : Proc.Roy.Soc.Ser.A 256,496-501)

112.Stocks,P.(1953).A study of the age curve for cancer of the stomach in connection with a theory of the cancer producing mechanism. Brit.J.Cancer 4,407-417

113.Sugiyama,H.(1961).Some statistical methodologies for epidemiological research of medical sciences. Bull.Intern.Statist.Inst.38,137-151

114.Sylvester,R.J.,Staquet,M.J.(1977).An application of decision theory to phase II clinical trials in cancer. In:Recent Advances in Cancer Treatment,Vol.3(H.J.Tagnon,M.J.Staquet eds.),pp.1-11. New York:Raven Press

115.Takâcs,L.(1954).Some investigations concerning recurrent stochastic processes of a certain type. Magyar Tud.Akad.Mat.Kutató Int.Közl. 3,115-128

116.Tallis,G.M.,Leppard,P.,O'Neil,T.J.,Sarfaty,G.(1979).The relationship of metastatic activity to primary tumor size with special reference to breast cancer. Austral.J.Statist.21,1-17

117.Tautu,P.(1973 a).Controlled Markov chains in medicine. Proc.4th Conf. Probability Theory (Brasov,Romania,1971),pp.461-469. Bucuresti:Ed. Academiei

118.Tautu,P.(1973 b).A note on the Savage's Sure-Thing Principle in treatment decisions. Seminar Probability Theory Oberwolfach,March 1973 (unpubl.paper,abstract Tagungsbericht 10/1973,p.12)

119.Tautu,P.(1976).Stochastic processes that act and interact. 9th European Meeting of Statisticians,Grenoble

120.Tautu,P.(1977).Processus patho-cliniques:dêtection et contrôle. In: Systems Science in Health Care (A.M.Coblentz,J.R.Walter eds.),pp. 153-165. London:Taylor & Francis

121.Tautu,P.(1978).Simulating the course of chronic diseases:Screening and therapeutic problems. Medizinische Informatik u.Statistik 8, 440-464

122.Tannock,I.(1978).Cell kinetics and chemotherapy:A critical review. Cancer Treatm.Rep.62,1117-1133

123.Thakur,A.K.,Rescigno,A.,Schafer,D.E.(1973).On the stochastic theory of compartments.II.Multi-compartment systems. Bull.Math.Biol.35, 263-271

124.Tolley,H.D.,Burdick,D.,Manton,K.G.,Stallard,E.(1978 a).A compartment model approach to the estimation of tumor incidence and growth: Investigation of a model of cancer latency. Biometrics 34,377-389

125.Tolley,H.D.,Manton,K.G.,Poss,S.S.(1978 b).A linear models application of competing risks to multiple causes of death. Biometrics 34,581-591

126.Tsakraklides,V.,Olson,P.,Kersey,J.H.,Good,R.A.(1974).Prognostic significance of the regional lymph node histology in cancer of the breast. Cancer 34,1259-1267

127.Tsiatis,A.A.(1978).An example of nonidentifiability in competing risks. Scand.Actuarial j.,pp.235-239

128.Underwood,J.C.E.(1972).A morphometric analysis of human breast carcinoma. Brit.J.Cancer 26,234-237

129.van Putten,L.M.(1974).Are cell kinetic data relevant for the design of tumour chemotherapy schedules ? Cell Tissue Kinet.7,493-504

130.Waalkes,T.P.,Tormey,D.C.(1978).Biologic markers and breast cancer. Seminars in Oncology 5,434-444

131.Weiner,D.,Purdue,P.(1977).A semi-Markov approach to stochastic compartmental models. Comm.Statist.Theory and Methods A6,1231-1243

132.Weiss,G.H.,Zelen,M.(1963).A stochastic model for the interpretation of clinical trials. Proc.Natl.Acad.Sci.USA 50,988-994

133.Weiss,G.H.,Zelen,M.(1965).A semi-Markov model for clinical trials. J.Appl.Probability 2,269-285

134.Werkheiser,W.C.,Grindey,G.B.,Moran,R.G.,Nichol,C.A.(1973).Mathematical simulation of the interaction of drugs that inhibit deoxyribonucleic acid biosynthesis. Molec.Pharmacol.9,320-329

135.Wheldon,T.E.(1978).Optimal control strategies in the radiotherapy of human cancer. In:Biomathematics and Cell Kinetics (A.J.Valleron, P.D.M.Macdonald eds.),pp.345-354. Amsterdam-New York-Oxford:Elsevier-North Holland Biomed.Press

136.White,C.C.(1974).A partially observed semi-Markov optimization problem. Ph.D.Thesis,Univ.of Michigan,Ann Arbor

137.White,C.C.(1976).Procedures for the solution of a finite-horizon partially observed,semi-Markov optimization problem. Operations Res.24,348-358

138.Wolfson,D.B.(1977).Limit theorems for sums of a sequence of random variables defined on a Markov chain. J.Appl.Probability 14,614-620

139.Woo,K.B.,Waalkes,T.P.,Ahmann,D.L.,Tormey,D.C.,Gehrke,C.W.,Oliverio, V.T.(1978).A quantitative approach to determining disease response during therapy using multiple biologic markers.Application to carcinoma of the breast. Cancer 41,1685-1703

140.Younge,P.A.(1965).The natural history of carcinoma in situ of the cervix uteri. J.Obstet.Gynecol.65,9-15

141.Zelen,M.(1966).Application of exponential models to problems in cancer research. J.Roy.Statist.Soc.Ser.A 129,368-398

142.Zwinggi,E.(1945).Versicherungsmathematik. Basel:Birkhäuser

Editor's Note

Because of illness the author had been prevented from presenting this paper in the meeting,and from submitting it for discussions. Taking into account the size of the paper as well as the scope of the present volume,the author agreed to its publication without the technical details brought together in the mentioned Appendices and Notes. The integral text represents a Technical Report,copies of which being available from the author upon request.

LIST OF PARTICIPANTS

Armbruster, Inge, Dipl.Inform.Med.	Institut für Medizinische Dokumentation, Statistik und Datenverarbeitung der Universität Heidelberg Im Neuenheimer Feld 325 6900 Heidelberg1
Baum, M. MD	King's College Hospital, Medical School Department of Surgery Denmark Hill, London SE5 8RX, England
Bässler, R., Prof.Dr.med.	Städtische Kliniken Fulda Pacelliallee 4, 6400 Fulda
Crowley, J.J. PhD	University of Wisconsin, Clinical Cancer Center, Departments of Human Oncology and Statistics 1120 West Johnson Street, Madison, WI 53715 USA
Eklund, G. Prof.	Medical Highschool, Göteborg Schweden
Fournier, D.v., Dr. med.	Universitätsfrauenklinik Voßstraße 9 6900 Heidelberg 1
Haybittle, J.L. PhD	Physics Department, Cambridge Health District, Addenbrooke's Hospital Hills Road, Cambridge CB2 2QQ, England
Immich, H., Prof.Dr.med.	Institut für Medizinische Dokumentation, Statistik und Datenverarbeitung der Universität Heidelberg Im Neuenheimer Feld 325,6900 Heidelberg 1
Kaae, S. MD	Aarhus Kommunehospital, Radiumstationen 8000 Aarhus, C. Denmark
Kalbfleisch, J. PhD	University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
Kay, R. PhD	Dept. of Probability and Statistics The University of Sheffield Sheffield S3 7RH, England
Keiding, N. Prof.	University of Copenhagen Universitetsparken 5 DK-2100 Copenhagen
Lagakos, S.W. PhD	Department of Biostatistics Harvard Univ. 677 Huntington Ave. Boston Mass.02115 USA

Mattson, Britta BA	The Oncologic Centre Radiumhemmet Karolinska Sjukhuset S-104 01 Stockholm, Schweden
Moeschberger, M.L. PhD	University of Missouri-Columbia, College of Arts and Science, Dept. of Statistics 222 Math.Sciences Bldg. Columbia, Missouri, 65211, USA
Müller, D.W.,Prof.Dr.	Institut für Angewandte Mathematik Im Neuenheimer Feld 294 6900 Heidelberg 1
Nelson, Sarah, J.	Sonderforschungsbereich 123 Im Neuenheimer Feld 294 6900 Heidelberg 1
Nissen-Meyer, R. MD	Oslo Kommune Aker Sykehis Oslo 5, Norwegen
Palmer, M.K. PhD	Christie Hospital & Holt Radium Institute Withington, Manchester M20 9BX, England
Peto, R. PhD	University of Oxford, Department of the Regios, Radcliff Infirmary Oxford OX2 6HE, England
Prentice, R. PhD	The Fred Hutchinson, Cancer Research Ctr. 1124 Columbia Street, Seattle, Washington 98104 USA
Ribeiro, G. MD	Christie Hospital % Holt Radium Institute Withington, Manchester M20 9BX, England
Roberts, Maureen MD	University of Edinburgh, Department of Clinical Surgery The Royal Infirmary Edinburgh EH3 9YW, Schottland
Scheurlen, H., Dr. med.	Institut für Medizinische Dokumentation, Statistik und Datenverarbeitung der Universität Heidelberg Im Neuenheimer Feld 325 6900 Heidelberg 1
Stewart, H., Dr.	Western General Hospital Dept. of Radiotherapy Crewe Road Edinburgh EH4 2XU Schottland
Tautu, P., Prof. Dr. med.	Deutsches Krebsforschungszentrum Im Neuenheimer Feld 280 6900 Heidelberg 1

Tsiatis, A.A., PhD	University of Wisconsin, Center for Health Sciences, Department of Preventive Medicine 504 Walnut Street (608)263-2880 Madison WI 53706 USA
Valagussa, Pinuccia BA	Istituto Nazionale Per Lo Studio E La Cura Dei Tumori via venezian, 1 20133 Milano/Italien
Wallgren, A. MD	The Oncologic Centre Radiumhemmet Karolinska Sjukhuset S-104 01 Stockholm Schweden
Wahrendorf, J.Dr.rer.nat.	Deutsches Krebsforschungszentrum Im Neuenheimer Feld 280 6900 Heidelberg 1
Weckesser, G. Dipl.Math.	Institut für Medizinische Dokumentation, Statistik und Datenverarbeitung der Universität Heidelberg Im Neuenheimer Feld 325 6900 Heidelberg 1