THE EVALUATION OF DURATIONS OF RESPONSE TO CANCER TREATMENTS

being a thesis submitted for the degree of

Doctor of Philosophy

in the Faculty of Science

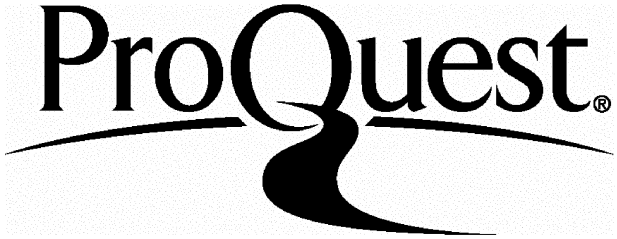
of the

University of London

in 1993

by

Walter Martin Gregory



To Mum

for setting such

a marvellous example

ABSTRACT

Abstract

The current methodology for scientific progress in clinical medicine is reviewed. Special attention is given to the randomised controlled trial, with particular emphasis on statistical considerations, especially the analysis of survival and response duration data.

The principles commonly employed in designing and improving treatments for cancer are reviewed, discussed, and demonstrated in practice by reference to progress made in the treatment of, among others, breast cancer and testicular teratoma. Improvements are shown to have concentrated on ways to circumvent resistance to therapy, and on exploiting the growth kinetics of tumours. However, quantitative information on either of these factors has been difficult to obtain, with the result that new trials are often designed on largely theoretical grounds, using principles that are unvalidated, and understood in a qualitative fashion only.

Two new mathematical models which seek to derive this quantitative information are presented, developed, and validated, and their assumptions discussed in detail. One population based model seeks to derive distributions of resistance and growth rate parameters for groups of patients from their durations of response to treatment. Applications are presented in acute leukaemia, breast cancer, Hodgkin's disease and multiple myeloma. A multivariate version of this model is presented and applied to help in the understanding and use of prognostic factors in breast cancer.

The second model, for individuals, uses sequential tumour volume measurements before each treatment. An application is given in lung cancer, where the volumes were measured by CT scan. Results from the model appear to indicate when changing or stopping treatment may be beneficial.

Application of these mathematical models often seems to generate new ideas for treatment, and leads to a better understanding of how the treatment may be working. They should enhance the conventional approaches, and hopefully enable research to proceed more rapidly and successfully.

TABLE OF CONTENTS:

List of tables

List of figures

Acknowledgements

I would like particularly to thank Ray Jackson, of the Clinical Operational Research Unit at University College London, for introducing me to mathematical modelling, and its applications in medical research. Both the theoretical and practical grounding I received from Ray in the field of operational research have been invaluable to me. At Ray's instigation I also had the good fortune to be attached to the ICRF Department of Medical Oncology at St. Bartholomew's Hospital, with Professors Jim Malpas and Andrew Lister. I spent a number of years there, learning how to analyse clinical data, and learning from the clinicians about the problems they experienced in understanding and interpreting their results, particularly their survival data for patients with cancer. The ideas that led to the development of the mathematical models reported in this thesis arose from this dual input of mathematical modelling combined with direct experience of clinical data.

I would like to thank the many people who have encouraged me with developing this work, and for their help in considering ideas and hypotheses with me until we came up with some that worked. In this area, I would particularly like to thank, in addition to the continuous help and support I received from Ray, Mike Richards, Maurice Slevin, and Bob Souhami. I'm also very grateful to Jim Burridge and Steve Gallivan and especially to Rodney Brooks for their technical advice with some of the mathematics, and to Rodney Reznek for carefully measuring so many small cell lung cancer tumour volumes.

I would also like to thank the numerous groups who have collaborated with me, both for providing their data, and often their encouragement and enthusiasm. In the latter category I would like particularly to mention Merrill Egorin and Leonard Reyno, with whom it has been a joy to work. The many others not already mentioned include Robert Rubens, Bob Leonard, Michael Baum, Michael Whitehouse, Lena Specht, John Rees and Richard Gray.

Introduction

The purpose of the work reported in this thesis was to develop and apply mathematical models that would be useful to clinicians treating cancer patients. As a result of working with cancer doctors for many years, it became clear to me that the analytic methods they employed, largely statistical in nature, while useful, did not answer all the questions they might. The observed results of clinical trials were often difficult to explain, and though the statistical tests employed might show that a particular treatment was better in terms of some particular endpoint such as patient survival (see chapter 2), they did not indicate why. The intention of using mathematical models was to generate useful hypotheses to help in explaining the results, and thus quicken the pace of research.

The usual outcome of a clinical trial is to conclude that one treatment is either the same or better than another treatment in some particular respect, e.g. patient survival, for the disease under study. The trial would usually be set up to answer the specific question as to whether one treatment was superior, and this question only. Often no attempt would be made to discover, in the trial design or the statistical analysis, *why* one treatment was superior. This difficult interpretive task would be left to the clinicians involved, who would then have to decide upon which treatments to use and evaluate next.

In parallel with this rather pragmatic approach to the development of more efficacious treatments, laboratory scientists were performing very detailed and

rational experiments designed to promote understanding of the basic biology of the diseases, and to model the manner of action of the different modes of treatment. However, the principles discovered in the laboratory often failed to produce useful results when applied to patients. It seemed that there were so many additional factors to be taken into account when treating a real patient, that the doctor was reduced to relying more on his clinical experience and intuition when deciding upon the treatment, rather than the desired logical and scientific explanations.

The mathematical modelling approach to problem solving appeared to provide a method for unravelling this complexity, and for deriving some of the fundamental assumptions upon which treatment strategies should be based. The essence of the modelling approach is the mathematical abstraction of the key elements of a complex process, with the aim of understanding the said process, predicting the results of altering certain features of the process, and therefore assisting with decisions concerning the process. It is worth noting that these models form the basis of Operational Research, which started in the 1940s and has met with success in areas where trials were infeasible, such as industrial and commercial environments. This philosophy appeared to have already met with dramatic, though limited, success in the clinical field through the models of Skipper *et al* (1964; 1967) and Goldie and Coldman (1979; 1982). Their models have had enormous influence on the design of clinical trials in cancer. However, these models embody principles, rather than having direct application to clinical trials and clinical data. There appeared to be enormous scope for more direct applications.

These models of Skipper (based on many experiments on animals in the 1960's) and Goldie and Coldman had been developed to try to explain observed trial results. However, the clinicians had little means of exploring the consequences of these hypotheses, and often relied merely on an intuitive feel for the results. It seemed that these various models needed to be quantified, and explored more thoroughly. The models needed to be tested, validated, and developed with the clinicians. The aim was to help explain their results more in terms of basic biology, to complement the more basic statistical outcome measures. Thus the clinician could perhaps be brought closer to the laboratory scientist, and the road to new and better treatments straightened.

The thesis thus begins with a brief history of the development of the clinical trial methodology, followed by a description of the current methods of analysing clinical research data. This latter review concentrates on the analysis of survival and response duration data, since most of the mathematical modelling applications are in this area. Some of the principles upon which treatment strategies have been developed are then described and discussed along with associated mathematical models. The thesis then proceeds to the new mathematical models that were designed and applied. These are described, along with their assumptions, and applications are given. A summary of the consequences of the modelling work in clinical applications is presented.

Chapter 1

CLINICAL TRIALS

Chapter 1

CLINICAL TRIALS

1.1 Definition

The essence of a clinical trial is the employment of a treatment on a **sample** of patients, to identify how best to treat patients in the future (Pocock, 1984). Clinical trials have thus been in existence for many centuries. However, in the last 40 years or so three new elements have been added to diminish or avoid bias, both in patient selection for studies, in allocation of patients to treatments, and in the use of historical data.

1.2 The Randomised Controlled Trial

Controlled studies, i.e. those in which a standard therapy is given to a group of patients similar to those being treated with the new approach, are generally considered necessary to avoid bias favouring a new treatment. This bias could arise because investigators may, perhaps subconsciously, select patients with a better prognosis for a new treatment, may not apply rigorous standards of assessment, may exclude patients who fail to benefit from the new treatment, or may prematurely abandon negative trials (Simes, 1986; Furberg & Morgan, 1987; Chalmers *et al.*, 1987; Green *et al.*, 1987). Furthermore, a positive psychological outlook to a new treatment may influence the patient, and the outcome.

Trials may also involve "blind" or "double blind" allocation of patients to treatments to avoid subjective bias in the approach of the clinician or the attitude of the patient.

As a result of investigation into agricultural experiments Fisher (1935) pioneered the method of **random** allocation. The intention was both to remove bias from the allocation procedures, and to avoid errors caused by some unknown factor being unevenly distributed between the groups.

Randomisation does not, of course, guarantee balance in other factors between the groups. However, the larger the trial, the less will be the chances of imbalance. To avoid imbalances in factors known to be important, trials can be stratified for these factors, with separate randomisations within the different strata. Despite this safeguard, imbalances can still occur, and even minor, non-significant imbalances in highly significant factors can confound the treatment comparison (Altman, 1985). Such factors should be adjusted for in the analysis even when stratification has been applied (Altman, 1985).

By employing **concurrent** controls, confounding time-related variables, such as changes in patient management and new supportive care techniques, can be avoided. Historical controls may bias the treatment comparison, and are best avoided. For example, consider the apparent improvements shown by early (phase II) trials of combination chemotherapy in multiple myeloma (Lee *et al.*, 1982). After some 18 randomised trials, spanning approximately 20 years, comparing combination chemotherapy with the previous standard therapy, namely melphalan and

prednisolone, it is still unclear whether the new treatments are an improvement (Gregory *et al.*, 1992). A study by Murphy *et al* (1986) provides an example of the unexpected influence of supportive care on survival in patients with acute myeloblastic leukaemia. Patients were randomised to receive leukocyte-poor blood components to prevent alloimmunisation or to receive normal blood components. On subsequent analysis several years later, those receiving the leukocyte-poor components appear to have an increased risk of relapse (Tucker *et al.*, 1989). Analysis of trials undertaken during these periods of changes in supportive care would be likely to exhibit major differences from current studies, due merely to the supportive care measures employed.

Trials may also require specification of the methods of analysis, rules for premature stopping of the trial (Pocock, 1984), an estimate of the numbers required to be treated, and rules for producing the final result of the trial. The latter often includes the specification of an end-date for the analysis, either after the entry of a specified number of patients, or at a particular time. These aspects are discussed more fully in chapter 2.

The Randomised Controlled Trial (RCT) methodology that resulted, combining these various elements and brought to clinical experiments largely due to the efforts of Sir Austin Bradford Hill (1952; 1962), now has a preeminent role in clinical science. In many circumstances RCTs are thought to provide the only definitive evidence of different treatment effects, and all but major breakthroughs require RCTs to establish the treatments involved as worthy. (An example given by Sir Austin Bradford Hill

where no trial was needed, was the use of streptomycin for tuberculous meningitis, where the disease was previously universally fatal (BMJ Editor, 1991)). A comprehensive outline of the philosophy and methodology can be found in Peto *et al* (1977a; 1977b). For critical reviews see Jackson (1985) and Birkhead and Jackson (1986). Using breast cancer as an exemplar, some of the benefits that have been derived from RCTs include the demonstration of the efficacy of adjuvant therapy (Bonadonna *et al.*, 1985; Bonadonna & Valagussa, 1987; Richards *et al.*, 1990) and the important negative finding that breast conservation gives similar survival results to mastectomy (Fisher *et al.*, 1985; Findlay *et al.*, 1985; VanderSchueren & VanDongen, 1988; Sarrazin *et al.*, 1989; Veronesi *et al.*, 1990). Over the past 20 years the RCT has come to be regarded as the *only* scientifically sound method of evaluating new treatments (Altman, 1984).

1.3 Ethics of the RCT

The aforementioned view should be treated with a degree of caution. Scientific developments can be so clear cut that it would be unethical to treat patients using older therapies which were known to be ineffective or greatly inferior (Armitage, 1975; Byar *et al.*, 1976). An example would be the treatment of diabetes with insulin (Black, 1979). Other ethical considerations may inhibit a clinician from performing an RCT (Armitage, 1984). It may be pointed out that there have been many instances in the past where clinicians have had strong beliefs that have subsequently turned out to be false. However, this can be guarded against by

alternative methods which do not involve subjecting patients to treatments which are very likely to be inferior. For instance, if a new treatment is expected to produce a *dramatic* improvement in survival, all new patients can be entered onto this treatment, with the intention of evaluating survival, with appropriate confidence limits, after the entry of some specified number of patients. If this survival rate is indeed a great improvement on the older methods, with confidence limits which easily exclude the previous survival figures, the use of inferior treatments will have been avoided. If the confidence limits include the previous survival figures, or leave some room for doubt, an RCT can still be performed.

Ethical dilemmas appear to be inherent in the RCT methodology. After all, the ultimate aim is to increase knowledge for the benefit of future patients, in an experiment which involves the treatment of individuals now. This may conflict with the clinician's responsibility to treat each patient in a way he or she considers optimum. There is a danger that the rules of the RCT become too rigid and uncompromising. There must come a point where the responsibility to the individual outweighs the possible future benefits to others.

Clinicians are not expected to undertake RCT's where they are confident that one treatment is better than the other(s). To undertake an RCT they must be in a state of uncertainty about the relative efficacy of the treatments. They are clearly expected to take all possible steps to avoid treating patients with inferior treatments. To this end, guidelines have been established for testing new treatments. Phase I studies constitute the first stage of such testing. Outside the field of cancer these are usually

conducted on healthy volunteers. For cancer drugs, which are likely to have considerable toxicity, they are usually conducted in patients for whom no treatment of proven benefit is available, largely with a view to establishing toxicity and safe dosages. Phase II studies examine activity of the drug/treatment, for drugs which have been through phase I testing. They are usually undertaken on patients who have failed primary therapy, and for whom no clearly beneficial alternative is available. It is usually only after encouraging phase II results that an RCT (phase III trial) is undertaken.

Looking at this problem from a different perspective, doctors must believe that two treatments give similar results, or at least be unsure of the outcome, in order to undertake an RCT. They have a duty to explain this to the patients, when giving them the choice of entry to an RCT. This 'informed consent' has gradually become an integral part of the RCT, and is probably a salutary exercise for the doctor, as well as being important for the patient. It ensures that doctors participating in the trial have given careful consideration to the subject.

Perhaps one additional ethical guideline should concern studies where a new treatment could not reasonably be imagined to fare worse, and is expected to be a considerable improvement on the current standard. If retrospective data is available concerning this treatment, this should be explored before an RCT is considered. New patients could all be entered onto the new treatment, while awaiting the results of retrospective analyses, including those from other institutions. As an example, consider the finding that the time of operation in premenopausal breast cancer

patients, in relation to phase of the menstrual cycle, affects their survival (Hrushesky *et al.*, 1989; Senie *et al.*, 1990; Badwe *et al.*, 1991). It would be possible to randomise patients to being operated on at supposedly "good" times in the cycle, versus random allocation as at present. However, doctors who were involved in the original retrospective data analysis and accept the results (a 30% survival difference at 10 years) would surely find this unethical. They felt that the policy needed to be changed immediately (M. Richards and I. Fentiman, personal communication); the only possible adverse consequences being a minor delay before surgery for some patients, which is thought highly unlikely to affect survival (Badwe *et al.*, 1991)). Meanwhile others who remain uncertain about the results can check the findings on their own retrospective data, where possible, and either confirm the findings, or cast sufficient doubt that a randomised trial would become a reasonable proposition.

1.4 History of the RCT

The first RCT to have a properly randomised control group was the Medical Research Council's pulmonary tuberculosis trial of 1948. At this stage, the tools for analysing clinical trials were imperfectly developed. Tests for comparing results where the outcome variable was a simple numerical value, such as blood pressure or weight, were straightforward (the t-test, or, for data not normally distributed, the Mann-Whitney non-parametric test (Armitage & Berry, 1987) could be used). However, the most common outcomes to be compared were survival times, or times to occurrence of some event, such as heart attack or relapse of disease. Since deaths (or events) would not have occurred in many patients, though information on the

length of follow-up would be available, the above tests could not be used. The data was always likely to be "incomplete" or censored in this way. A major breakthrough, therefore, in this field was the derivation of tests for comparing censored survival data, developed by Mantel (1966), and discussed more fully by Peto and Peto (1972).

In the design stage of an RCT, the clinician must decide what size of difference he is looking for, and thus how many patients are required to have a reasonable chance of detecting such a difference. Statistical methods have been derived for generating such numbers (see, for example, Freedman (1982)), and are discussed more fully in the next chapter. However, practical considerations have meant that trials have frequently been conducted with insufficient numbers to detect likely differences. Clinicians have tended to be over optimistic in their expectations of new treatments, and it is often felt that having some trial underway, even if it only has a slight chance of detecting a difference, is better than not running a trial at all. One additional, often neglected, reason for the multiplicity of small studies is the perceived need for (first) authorship on papers. Research institutions are measured and evaluated to a large extent on their publications, and clinicians are likely to enhance their future employment prospects considerably by publishing research papers. Multi-centre studies decrease the likelihood of such authorship, and in this way discourage participation.

There are two possible approaches to this problem of insufficient numbers of patients being recruited to trials. Statisticians have appealed for large multi-centre trials to

be undertaken answering some of these key questions (Yusuf *et al.*, 1984). However, such trials are difficult to organise, and have had only limited, though sometimes unequivocal, success (for example, ISIS-2 (Second International Study of Infarct Survival, 1988)). Large organisations appear to be necessary to run such trials, and several have been formed, particularly in the USA. Examples include the South West Oncology Group (SWOG), the Eastern Co-operative Oncology Group (ECOG), the European Organisation for Research on Treatment of Cancer (EORTC), The Ludwig group, the National Surgical Adjuvant Breast and Bowel Project (NSABP) and the 'Nolvadex' Adjuvant Trial Organisation (NATO). An alternative approach, which will now be discussed, has been the development of meta-analyses or "overviews".

1.5 The Overview Methodology

As a result of the above considerations, in many cancers many slightly different trials have been undertaken addressing the same, or very similar, questions. Examples include the evaluation of tamoxifen and adjuvant chemotherapy in early breast cancer (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 1992), the role of multi-drug combination chemotherapy in breast (A'Hern *et al.*, 1988) and ovarian cancer (Slevin, 1986), whether combination chemotherapy is superior to melphalan plus prednisolone in multiple myeloma (Gregory *et al.*, 1992), and whether the addition of primary chemotherapy to radiotherapy improves the survival of early stage Hodgkin's disease (see the reviews and the international workshop results in Somers *et al* (1990)).

The clinical trial methodology has therefore recently been extended to include the combined analysis of trials addressing similar questions; so called "overviews" (for examples, see Yusuf *et al* (1985), and Canner (1987)). The overview methodology combines information from different trials, in a similar way to that in which the RCT itself combines the results for individuals treated in the same manner. A weighting is attached to each trial's result based on the number of patients entered, and an overall, weight adjusted, mean expected difference is produced, with associated p-value based on the summed log-rank test statistics of the individual trials (see chapter 2 for details of the log-rank test, and Peto (1987) for the "overview" methodology).

This departure from the more strict rules described earlier is to be welcomed. It is similar to the stratification methods outlined above, but recognises that trials with marginal differences, but addressing similar general questions, can produce useful results. Differences between the trials can be quite marked when their data is pooled. For example, breast cancer overviews have combined all single agent chemotherapy versus combination chemotherapy trials (Early Breast Cancer Trialists' Collaborative Group, 1992).

1.6 Difficulties with the RCT philosophy

Ethical dilemmas and analytical problems with the RCT philosophy have already been discussed. The RCT philosophy raises other difficulties. The degree of generalisation possible from an RCT is problematic. Since the RCT rules are

designed to ensure that the same conditions exist for patients in each arm of the trial, other trials at different centres may address the same questions under slightly different conditions, and produce different answers. As an example, consider the many trials comparing intensive combination chemotherapy with melphalan and prednisolone in multiple myeloma (Gregory *et al.*, 1992). Two of the 18 published randomised trials showed highly significant differences ($p=0.004$, and $p<.02$), in opposite directions! When analysed using the overview methodology there was no overall difference between the treatments, but these two trials had 99% confidence limits which did not include a 'no difference' result, and thus conflicted with this finding (Gregory *et al.*, 1992). These results have caused considerable controversy, with proponents of both viewpoints arguing vehemently for their beliefs (for example, see Bergsagel (1989) and Lee (1984)).

The "overview" analyses have to make some assumptions to deal with this problem. Two approaches have been suggested. The most common is to assume that all trials are estimating the same true *fixed effect* of treatment (Yusuf *et al.*, 1985). The trials are then handled as strata, as discussed above. The other approach is to recognise that trials may be heterogeneous in the sense of having differing true effects. These unknown effects are then represented by a *random effects* model (DerSimonian & Laird, 1986; Armitage & Parmar, 1986). The former method seems particularly open to question if there is significant heterogeneity between centres, sufficient to cast doubt on the results, while the latter has distinct problems in handling the distribution of the random effects (Pocock & Hughes, 1990). A pragmatic solution would seem to be only to trust the overview results when there isn't a significant

heterogeneity effect, in which case the two methods are likely to give similar results (Pocock & Hughes, 1990). This has much in common with the mathematical modelling approach, in that common patterns in the different centres are being sought. The modelling approach would suggest further analysis or modelling where there was significant heterogeneity, to discover the reasons for this heterogeneity.

Similar problems may also arise in the standards of measurement, and thus classification. For instance, when comparing results in Hodgkin's disease from St. Bartholomew's hospital, in London, with the Christie hospital, Manchester, different proportions of patients appeared to present with the various histological subtypes in the two centres (Wagstaff *et al.*, 1988). A recent review of the classification of Hodgkin's and non-Hodgkin's disease patients by a panel of pathologists, who examined the slides separately without prior knowledge or consultation, showed interrater agreements which fell as low as 60% for some pathological subgroups (Hanby *et al.*, 1992). Trials where stratification by such factors has been employed would therefore be open to question.

Another philosophical dilemma concerns the general mechanisms by which medical science advances. Would advances be more likely as a result of many small different trials being undertaken, looking for large differences; or a few large RCT's looking for small differences? There is probably room for both, with highly specialised research hospitals undertaking small innovative studies, while more widely based large multi-centre groups address questions perhaps raised by the smaller centres.

There is a view that large treatment differences are hardly ever found, particularly as time goes on and more possibilities have been exhausted (Yusuf *et al.*, 1984). The appeal for large multi centre studies as the only reasonable method of advancement derives from this belief. However, it is difficult to justify this latter stance. Consider as a counter example the recent study by Slevin *et al* (1989) on only 39 patients in small cell lung cancer, where an improvement in response rate of 74% (from 10% to 84%), and a consequent improvement in survival, was found merely by altering the scheduling of one drug. This study was also randomised, implying that small randomised studies can be viable and worthwhile.

A further example can be found in the treatment of testicular teratoma. Cure rates for patients with stage IV metastatic disease rose dramatically from < 10% in the 1960's (Mackenzie, 1966; McElwain & Peckham, 1974) to 80-90% in the late 1970's and 1980's, when high doses of new and more effective platinum based regimens became available (Newlands *et al.*, 1983; Peckham *et al.*, 1983; Vugrin *et al.*, 1983; Oliver, 1986). RCTs were not necessary for these studies; large obvious improvements were found in single centres. Having made this leap in response rate and survival, many subsequent studies in teratoma have been randomised, in an attempt to find the best drugs and schedules to combine with the platinum based drugs. The variety and type of study undertaken is clearly related to the size of the expected (or hoped for) differences. Thus it seems that there is a role for both the large multi-centre study, and the small, innovative, single research institute trial.

Although the RCT is a powerful tool for deciding which of several treatments is better (in terms of some outcome measure), it is not, of course, of any help in understanding, per se, the mechanisms by which the treatments work. Previous RCTs may provide information which the clinician may use in trying to understand these mechanisms, but these are likely to be only one part of a design process involving many facets. An understanding of the biology of the disease, the pharmacokinetic action of the drugs, the resistance and cross resistance profiles of the drugs, the immunology of the host, the toxicities of the treatments, and a multitude of other factors all play a role in the design.

Because of the plenitude of such factors, and the difficulties in understanding their actions and interactions, mathematical models have been developed to try to enhance the understanding of these processes, and to provide some principles to guide in the design of new treatments. Indeed, these have sometimes become embodied as clinical dogma, being used repeatedly over many years as a fundamental aspect of the design of new regimens. Consider, for example, the hypothesis developed by Skipper, that a treatment will kill a constant *fraction* of the disease on each administration (Skipper *et al.*, 1964; 1967), or the hypothesis of Goldie and Coldman that two equally effective drug combinations which have some degree of non-cross-resistance will be most effective administered in an alternating fashion (Goldie *et al.*, 1982). To put the RCT in perspective therefore, it is a tool to be used towards the end of the research process to test or verify previously derived hypotheses. This role is, of course, vital; many, if not most, new treatments do not improve on existing standards and therefore fail the RCT test.

Recently, an alternative methodology for arriving at the optimum dosages and scheduling of drugs in combination has been suggested. It was observed that with the number and variety of drugs and doses available in many cancers, the possibilities for treatment were so great that the chances of nearly optimum doses and combinations being found by the use of RCTs was remote. The essence of the new approach is to treat small cohorts of patients with different doses and schedules, with a rule based on mathematical hill climbing algorithms to determine the next dose and schedule (Berenbaum, 1990). This process is continued until an optimum result is reached, with response or survival no longer improving. This is almost the antithesis of the RCT approach; indeed it is explicitly stated that treating too many patients with the same doses will reduce the effectiveness of the procedure. This methodology has, to date, only been used in animal studies. However, it yielded survival improvements never previously observed, despite employing conventional doses of drugs which have been tested extensively over many years.

1.7 Conclusions

The randomised controlled trial methodology has clearly been a great step forward in bringing more scientific rigour to the field of clinical research. However, there has to some extent been a swing away from the old extreme of too much subjective interpretation, to a new, opposite, extreme of too many rules, and an inflexible approach. This is being relaxed somewhat with the "overview" approach, and other valuable methods of analysis such as that described by Berenbaum (1990). The

mathematical modelling approach used in this thesis should continue this trend, while maintaining the necessary scientific rigour.

Chapter 2

OUTCOME MEASURES AND CONVENTIONAL METHODS OF ANALYSING SURVIVAL DATA

Chapter 2

OUTCOME MEASURES AND CONVENTIONAL METHODS OF ANALYSING SURVIVAL DATA

2.1 Introduction

To complement the principles outlined in the previous chapter for designing a clinical trial, a clear and unambiguous protocol is important. The rules should include eligibility criteria, instructions as to what operations, drugs, and other clinical measures to use at particular times during the patient's treatment, and what supportive care measures are to be employed. The outcome measure should also be defined. This is related to the desired size of the cohort of patients to be treated - which should also be specified - and will now be examined in greater detail.

2.2 Outcome measures

2.2.1 Definition

A RCT will be designed to compare some particular outcome(s). This can be the survival time of the patient, the patient's response to treatment, or any of a large number of other measured endpoints such as the patient's blood sugar level, the quality of the patient's vision, their haemoglobin level or their quality of life.

The statistician becomes involved in the clinical trial process at several stages. This involvement is fundamentally concerned with this outcome measure. To determine how many patients should be entered into the trial, it is necessary to have some *a priori* knowledge of the outcome measure. This is because the distribution and variability of this measure will limit the usefulness of the results. For example, suppose the aim of a particular trial was to increase the patient's weight. If it was thought that every patient would gain 1 stone in weight as a result of the treatment, and the first few patients all gained 1 stone, as predicted, the trial could be completed rapidly, with a small number of patients. However, a more likely scenario would be that some patients would gain a lot of weight, others gain some weight, others lose weight or remain the same weight, with the average weight gain being 1 stone. If this variability were considerable, it would clearly be very difficult to evaluate the results from only a few patients. Thus determination of the number of patients needed to be entered into the trial requires an estimate (or, in the lack of any available data, a guess) of the trial's results in terms both of the outcome measure and it's variability, before the trial has started.

2.2.2 Testing for differences

To evaluate the trial's results, outcome measures need to be assessed, quantified, and compared. Appropriate statistical methods for comparing different outcome measures between two or more treatments have been determined for all the outcome measures mentioned above. However, this has only been done for a simple comparison of the treatments, i.e. an answer to the question is the outcome of one treatment different

from that of another. Alternative, more structured outcome measures will be described later.

For a simple numerical outcome measure, such as haemoglobin, where the outcome measure, or a simple transformation of the outcome measure, might be expected to be normally distributed, the trial results can be compared using either the t-test when comparing two groups, or using analysis of variance methods (Armitage & Berry, 1987) for more than two groups. Some factors measurable before the start of the trial may also be related to the outcome measure (they are sometimes called prognostic factors), and may be imbalanced in the different arms of the trial (for example, one arm may contain more older patients). These factors, often called nuisance factors, can usually be adjusted for using linear regression methods (Armitage & Berry, 1987). Similar non-parametric methods are available for outcome measures which are not normally distributed, for example, the Mann Whitney U test, and Kruskal Wallis analysis of variance.

For a simple categorical outcome measure (i.e. when the outcome is one alternative from a set of possibilities, e.g. response to treatment, failure, or death) the trial results can be evaluated using the chi-squared test (Mantel & Haenszel, 1959). Multivariate logistic regression can be used to adjust for nuisance factors (Engelman, 1985).

2.2.3 Power calculations and trial size

Before describing methods for determining how many patients need to be entered into such trials, it is necessary to describe and define two concepts. There will usually be two possible outcomes for a trial comparing two treatments, namely that one treatment produces better results or that the treatments produce the same results. Whichever of these conclusions is reached, there remains a chance that the conclusion is false. Thus it is possible to have a false positive result, or a false negative result (these are often called type I and type II errors respectively). The aim of determining the numbers required to be entered into the trial, is to limit the chances of false positives and false negatives to some prescribed probabilities. The relevant computations are often called power calculations, since they evaluate how powerful the trial will be in detecting or excluding differences, given the numbers of patients entered. The most commonly chosen false positive and false negative probabilities are .05 and .1 respectively, but can depend heavily on the aims of the trial. Thus the researcher often has to accept that although the trial may have a positive outcome, there will still be a 5% chance that this outcome is false, or that, although the trial has a negative outcome, there will still be a 10% chance that this outcome is false, and that there is really a difference between the treatments.

The numbers required for simple numerical or categorical outcome measures as described above, have been determined and are described in Pocock (1984). The situation is more complicated for the analysis of trials where the outcome measure is the survival or some other event time for the patient. To compare two groups of

survival times, the most commonly used statistical test is the log-rank test (Mantel, 1966; Peto *et al.*, 1977b). Power calculations for determining trial size based on the log-rank test are described by Freedman (1982). These are based on examination of the distribution of the log-rank test statistic, which can be shown to be approximately normally distributed. Since one major thrust of the new mathematical modelling methods involves alternative methods of analysing survival and duration of response data, the conventional methods for such analysis will now be described and discussed in more detail.

2.3 Conventional methods of analysing survival or event time data

2.3.1 Survival times

If the patient has died, the survival time is the time from their entry into the trial (usually the start of treatment) until death. If the patient is still alive, the survival time is the time from entry into the trial until the time the patient was last seen. In this latter case the patient's survival time is said to be 'censored' at this point. This censoring aspect of survival data complicates the analysis, and has entailed the development of new statistical methods and tests.

Analyses are commonly performed on time intervals other than survival, e.g. for patients whose disease is no longer discernible after treatment, the time taken for the disease to reappear (this is usually called relapse) is frequently analysed. This time

is referred to as the duration of response or the duration of remission. The methods of analysis which will be described in the remainder of this chapter apply equally to both survival data and other time intervals, though survival time will be used as the exemplar.

2.3.2 Actuarial survival curves

The accepted method of presenting survival data is to plot the percentage of patients alive as time increases from entry into the trial. This cannot be evaluated simply as a proportion of the number of patients left at each time because of the patients who are censored.

Kaplan and Meier (1958) therefore produced the following estimate of this survival percentage, (often called the actuarial survival method, since it is commonly employed by actuaries):

Consider any time t, after the start of treatment. Suppose n patients are still at risk of dying at this time (the others have already died, or are censored before this time). Suppose d_t patients died on day t. Then the probability, P(t+1), of surviving to the next day, given that the patient is alive at t, is

$$P(t+1) = \frac{n-d_t}{n}$$

$P(0)=1$ so the % surviving up to time t is:

$$\prod_{i=1}^{t} P(i)$$

This method can be considered intuitively as representing the chances of the patient leaving an interval that they have entered.

This definition means that the censored data is used to the full, i.e. probabilities before a particular censored time depend on that censored time; probabilities after this time do not. This makes intuitive sense, since the fact that a patient has survived 20 days without dying provides information about the chances at 10 days though it does not give comparable information about the patients' chances at 30 days.

In practice it is not necessary to evaluate the survival percentage on every day. It will only fall on days when there are deaths, so only those days need to be considered.

As an example, consider the following survival times (in days):

12 27 53 61* 69 75 113* 159

where *'s indicate censored times.

The survival %'s are:

time	12	27	53	61*	69	75	113*	159
%	87.5	75	62.5	62.5	46.9	31.3	31.3	0

and the actuarial curve therefore looks like

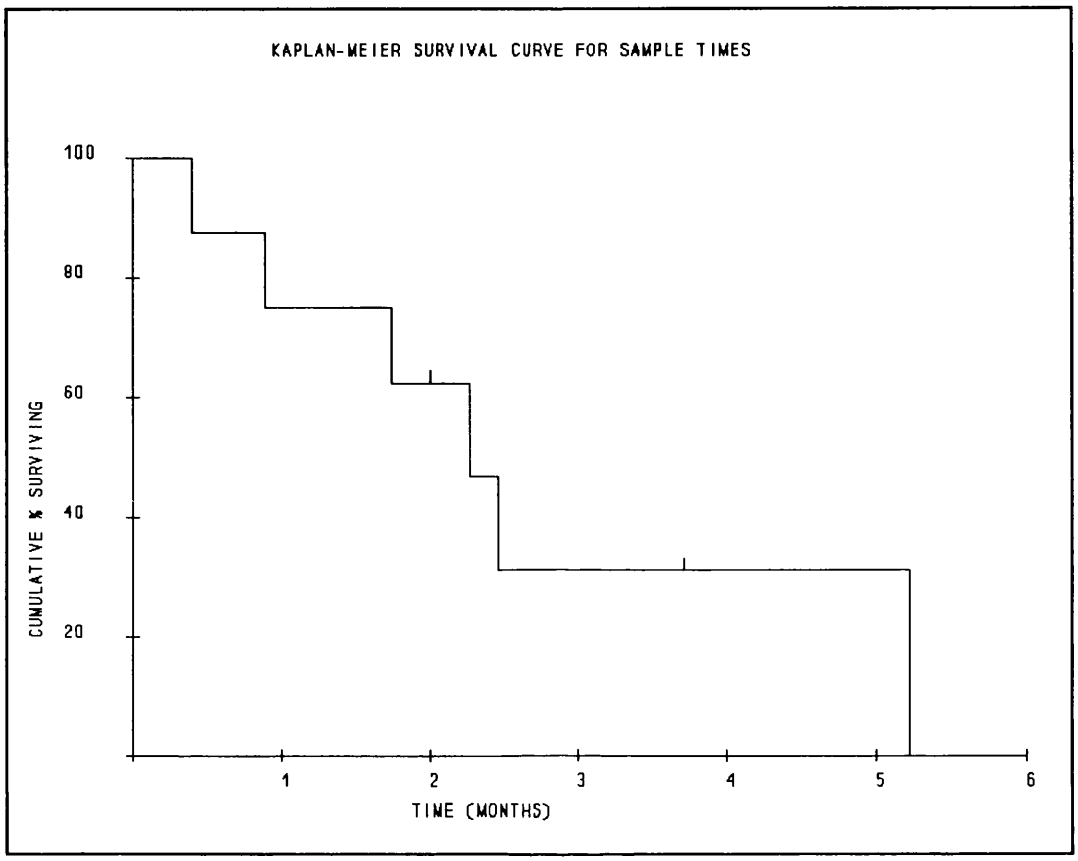


Figure 2.1 Actuarial survival curve for the sample times shown above.

The blips (upwards pointing marks) on the curve represent the censored times, and display the patient's follow-up times at a glance. This is useful for evaluating the precision of the survival estimates (confidence limits could also be used to this end, and will be described shortly). To see this more clearly consider the following two curves, taken from the same trial analysed during entry to the trial, and analysed again several years later.

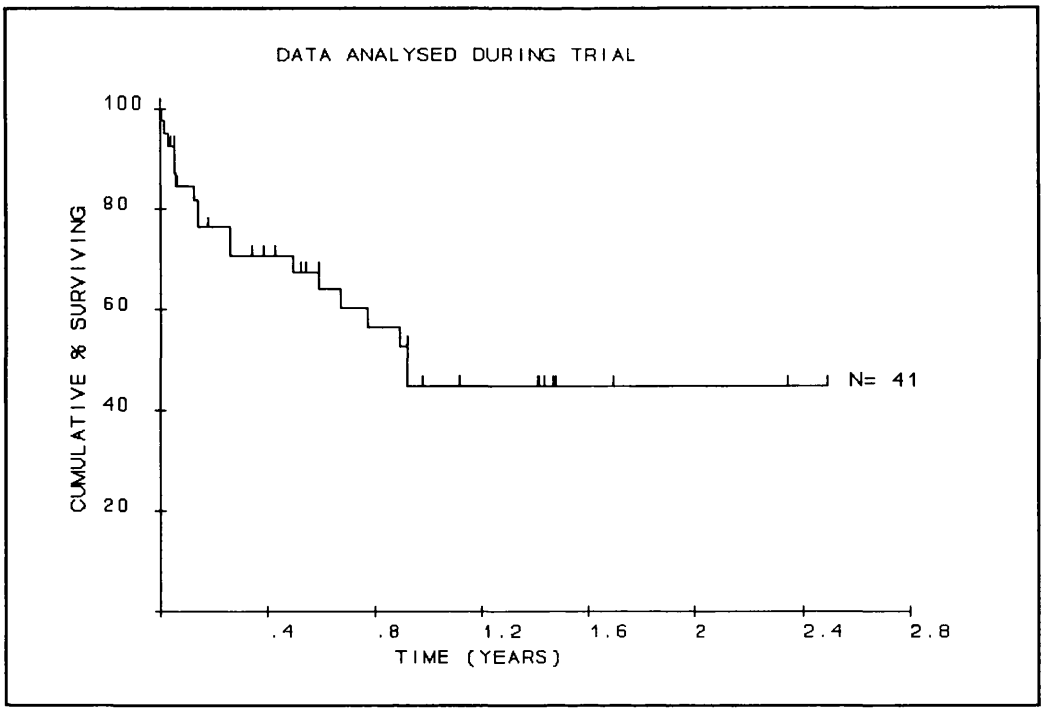


Figure 2.2 Analysis of survival during a trial

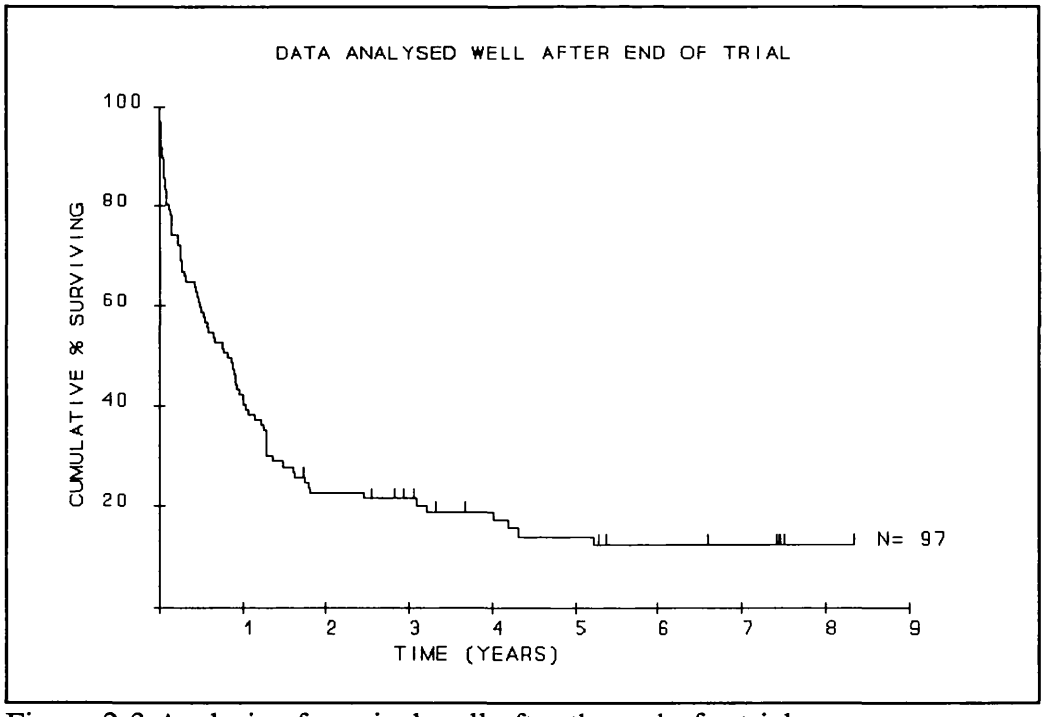


Figure 2.3 Analysis of survival well after the end of a trial

The censored points on the latter curve occur after the vast majority of the deaths. Thus it can be confidently asserted that this curve will not alter greatly. In particular, it cannot alter in the period before 18 months, since there are no patients at risk of dying during this period. In contrast, the earlier curve contained patients at risk throughout the whole time interval, and indeed the curve changed considerably after longer follow-up.

Two or more survival curves can be plotted together to compare the survival of different groups of patients. A statistical comparison between the curves is necessary to test whether the differences observed could have occurred by chance, or whether there is likely to be a real difference in survival between the two treatments.

To test for such differences some assumptions must be made about the distributions of the survival times. The most commonly adopted assumption concerns the hazard rates in the two or more groups. The hazard rate is the instantaneous risk of death at any given time. For many trials and treatments this will change over time, often diminishing as the period from initiation of treatment increases. The hazard rate can be thought of as the slope of the survival curve. Most of the statistical methods for analysing survival data assume that the hazards of the groups to be compared are proportional, i.e. the hazard in one group is the same linear function of that in any of the others, at all times.

Given the proportional hazards assumption a statistical test known as the log-rank test (Mantel, 1966; Peto *et al.*, 1977b) is the most efficient test for comparing

differences in survival times between groups of patients (Peto & Peto, 1972). This will now be described.

2.3.3 The Log-Rank Test

The principle of the test is as follows: if there is no difference in survival between two groups of patients, then it would be expected that, at any time, a patient would have the same chance of dying in one group as in the other. We would therefore expect the number of deaths at any time to be distributed between the two groups in proportion to the numbers at risk.

Suppose that at time t, n_1 patients are still at risk in group 1, and n_2 patients are still at risk in group 2, and that d_t patients die on day t. Then (Armitage & Berry, 1987)

$$E(d_{1t}) = d_t \cdot \frac{n_1}{n_1+n_2} \qquad E(d_{2t}) = d_t \cdot \frac{n_2}{n_1+n_2}$$

$$\text{var}(d_{1t}) = \text{var}(d_{2t}) = \frac{d_t (n_1+n_2-d_t)n_1n_2}{(n_1+n_2)^2(n_1+n_2-1)}$$

where $E(d_{it})$ and $\text{var}(d_{it})$ = expected number and variance of deaths in group i on day t (i=1,2).

The expected deaths can be compared with the observed deaths as follows:

Let $E_1 = \sum_k E(d_{1t})$, $E_2 = \sum_k E(d_{2t})$ $V = \sum_k \text{var}(d_{1t})$

where the sum is over all k deaths. Let

O_1 = number of deaths in group 1

O_2 = number of deaths in group 2

Peto *et al* (1977b) assume that O_1-E_1 (and O_2-E_2) approximates to a normal distribution. Then a test statistic for equivalence of the death rates in the two groups is

$$\chi_1^2 = \frac{(O_1-E_1)^2}{V}$$

An alternative and simpler approximation (Peto *et al*, 1977b) to the variance, V, of this normal distribution is

$$\frac{1}{\frac{1}{E_1} + \frac{1}{E_2}}$$

and the statistic $(O_1-E_1)^2/V$ is then compared to a χ_1^2 distribution. The more familiar formula is then easily derived, since

$$\frac{(O_1-E_1)^2}{V} = \frac{(O_1-E_1)^2}{\frac{1}{E_1} + \frac{1}{E_2}} = \frac{(O_1-E_1)^2 E_1 E_2}{(E_1+E_2)}$$

$$= \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2}$$

This can be easily extended to n groups:

$$\chi_n^2 = \sum_{i=1}^{n} (O_i-E_i)^2/E_i$$

It should be remembered that this is a heterogeneity test - i.e. it merely tests to see if all n curves could have arisen from the same underlying distribution. It does not test for individual differences between curves.

A variant of the log-rank test can be applied where there may be a trend for improving (or worsening) prognosis. The trend test is defined as follows:

Let each subgroup be given a number, n, starting at 1, and increasing by 1 for each group.

Calculate O's and E's for all n groups. For each group let

$A = n(O-E)$

$B = nE$

$C = n^2E$

Then the trend statistic, T,

$$= \frac{(\Sigma A)^2}{V}$$

where $V = \Sigma C - (\Sigma B)^2/\Sigma E$

Then $T \approx \chi_1^2$

2.3.4 Standard errors and confidence limits on survival curves:

If the value on a survival curve at a particular time is P, then an approximate standard error for P at that time is

$$P\sqrt{((1-P)/N)}$$

where N is the number of patients still at risk at that time (Peto *et al.*, 1977b).

For example, if P=0.5 at 1 year, and there are 80 patients still at risk at this time, then SE(P) at 1 year = 0.04. Suppose that the curve has dropped to 0.2 at 3 years, and there are now only 10 patients at risk, then SE(P) at 3 years = 0.06.

These estimates can be used to derive confidence intervals for a survival curve at any desired time. More accurate confidence interval estimates are provided by Simon (1986). His method involves deriving an 'effective sample size', n, at the desired time, accounting for the censored times. The confidence interval is then derived from the actuarial percentage, p, at the time in question (using the formula for calculating confidence intervals for a response rate), and equals

$$\{ p + A/2 \pm Z\sqrt{p(1-p)+A/4n} \}/(1-A)$$

where $A=Z^2/n$, Z being a standardised normal deviate, (so Z=1.96 for a 95% confidence interval, 1.645 for a 90% confidence interval etc.). The effective sample size, n, is given by

$$n = (1-p)/\{ p \sum_{t_i<T} d_i/(r_i-d_i)r_i \}$$

where p, as mentioned, is the actuarial probability of survival beyond time T, t_i denotes the i^{th} smallest distinct time of death, r_i is the number of patients at risk just before t_i, and d_i denotes the number who die at t_i.

Chapter 3

THE PRINCIPLES OF CANCER THERAPY

Chapter 3

THE PRINCIPLES OF CANCER THERAPY

3.1 Evolution of the principles

The guiding principles on which much of modern cancer therapy is based originate with experiments undertaken by Skipper and colleagues in the 1960's (Skipper *et al.*, 1964; 1967). As a result of his investigations of experimental tumour systems, Skipper concluded that a given dose or course of (chemo)therapy will kill a constant *fraction* of the cell population, rather than a constant number of cells. He went on to consider the consequences of tumours having different growth rates, and concluded that chemotherapy gave a greater fractional cell-kill to more rapidly growing tumours, and was more likely to be curative in such cases (Skipper & Perry, 1970). It also appeared that larger tumours generally grew more slowly, and that the lack of responsiveness of such tumours was related to this slower rate of proliferation (Shackney, 1970; Steel *et al.*, 1976).

These results apply to radiotherapy as well as chemotherapy (see for instance, Okumura *et al* (1977)), and have now been explored in considerable detail. Larger tumours are generally considered to grow more slowly because, although early (sub-clinical) tumour growth follows an exponential pattern, tumour growth slows down as the tumour increases in size, presumably due to problems of nutrient supply, approaching a maximum volume. Laird (1964) proposed a Gompertzian

function to describe this growth, where the (exponential) growth rate also declines exponentially, resulting in such a maximum volume. The Gompertzian model appears to fit the experimental data well for a variety of tumours and growth rates (Sullivan & Salmon, 1972; Demicheli, 1980; Akanuma, 1983; Pearlman, 1983). Thus the unresponsiveness of a tumour may be highly dependent on the point in its growth curve at which therapy is initiated.

It is interesting to note that with a Gompertzian growth curve, effective treatment of large indolent tumours can result in smaller, more rapidly dividing tumours, with new possibilities for treatment. Norton and Simon (1977; 1986) suggested that initial induction treatments, at modest doses, could be used merely as a method of reducing the tumour to a size where it grew more rapidly. At this point, intensive therapy could be initiated in an attempt at cure.

Following on from these early principles, Goldie and Coldman (1979) examined the success or failure of therapy from a different point of view, namely the presence or the acquisition of resistance. They demonstrated that if there was a constant rate of mutation towards resistance (i.e. if every time a cell divided it had the same chance of mutating to become resistant), then there would be a critical and *short* period in the tumour's history when the chances of cure dropped from 1 to zero. (Essentially there would be a critical 'mass' of tumour cells beyond which the chances of a mutation occurring would be very high). Examples were given where this occurred in a 1-log range, e.g. from 10^7 cells to 10^8 cells. Thus it would be vital to treat during, or preferably before, this period. Of course, the more rapidly the tumour

grew, the more important it would be to treat early, since the tumour would be quickly progressing to the point where resistant mutants would inevitably arise.

To summarise these results, these hypotheses related to tumour growth and resistance suggest that the intensity, frequency and duration of therapy should be matched to the tumour's growth rate and to the point reached in the tumour's growth curve (essentially the tumour size). The more rapid the growth rate, the more intensive, frequent and short lived should be the therapy. For slowly growing tumours, longer durations of therapy are likely to be necessary, probably at reduced doses. The aim, in these latter cases, should be to eliminate the dividing cells, which may be a small fraction of the tumour. Higher doses are likely to be unproductive, killing few extra cells, and possibly compromising later therapy due to factors such as toxicity and acquisition of resistance.In addition, therapy should be given as early as possible, to maximise the chances of cure, by treating before resistant mutants have arisen. This will be especially critical in rapidly growing tumours, where a short delay could allow a large increase in tumour size, and a consequent severe reduction in the chances of cure.

3.2 The principles in practice

These principles can be applied not only to the choice of treatments for different cancer types, but to the choice of treatment for different individuals with the same

type of cancer based on the growth rates of their tumours. A number of examples will be given to demonstrate these points.

3.2.1. Application to particular cancers

Consider firstly, a very rapidly growing malignancy like testicular teratoma, where the therapist should aim to administer the maximum dose in the minimum time. Cure rates for patients with stage IV metastatic disease have increased dramatically, rising from < 10% in the 1960's when treatment was spread over 2 or more years (Mackenzie, 1966; McElwain & Peckham, 1974) to 80-90% in the late 1970's and 1980's, when high doses of new and more effective platinum based regimens, typically administered for only two or three courses, were introduced (Newlands *et al.*, 1983; Peckham *et al.*, 1983; Vugrin *et al.*, 1983; Oliver, 1986).

In contrast, consider the treatment of gastric cancer. Initial adjuvant studies conducted in Japan comprised moderate doses of Mitomycin C spread over a period of 5 weeks, and achieved modest but definite success (Imanaga & Nakazato, 1977; Nakajima *et al.*, 1978). With the popularity of intensive scheduling, subsequent trials administered larger doses over shorter time periods, including treatment concurrent with surgery (Imanaga & Nakazato, 1977). These trials not only failed to improve on the earlier results, but any beneficial effect appeared to have been lost. Examination of times taken for early untreated gastric cancer patients to progress to late stage (Tsukuma *et al.*, 1983) demonstrate that gastric cancer has a slow growth rate, and thus longer, sustained therapy may be necessary. It may be that five weeks

of therapy is inadequate, and that further benefit would be gained by even more prolonged therapy.

In many cancers, trials addressing questions of duration of treatment are now being undertaken. For instance, in Wilms tumour as a result of a series of trials over the last 21 years, the standard duration of therapy has been reduced from 15 months to 10 weeks (D'Angio *et al.*, 1976; 1981; 1989). Similar trials and comparisons have also been undertaken in Hodgkin's disease (Young *et al.*, 1973; Medical Research Council's Working Party on Lymphomas, 1979; De Vita *et al.*, 1980), non-Hodgkin's lymphomas (Connors & Klimo, 1988) and Leukaemia (Bell *et al.*, 1982; Vaughan *et al.*, 1984) among others, supporting the use of short intensive induction regimes without maintenance therapy in these diseases.

In early trials of adjuvant therapy for patients with breast cancer, the typical duration of treatment was 12 months (Bonadonna *et al.*, 1985; Bonadonna & Valagussa, 1987; Richards *et al.*, 1990). It has subsequently been shown that equivalent results can be achieved with only 6 months treatment (Bonadonna, 1985). However, trials in which prolonged (6 months or more) treatment was compared with single course (perioperative) treatment have demonstrated that prolonged treatment is more effective (EBCTCG, 1992). It is still unclear whether, for example, 3 months treatment would be as effective as 6 months treatment. Again it is apparent that breast cancer is a relatively slowly growing tumour. For instance, using the monoclonal antibody Ki67, which is reported to stain cells not in the G0 phase of the cell-cycle and to give reliable estimates for the growth rates of a number of

tumours, less than 20% of breast cancer cells are stained on average (Gerdes *et al.*, 1986; Barnard *et al.*, 1987). This compares with an average of greater than 50% in, for example, high grade non-Hodgkins lymphomas (Gerdes *et al.*, 1984). Again, moderate dose, longer duration treatment seems to be required.

Minimising the duration of therapy may, on the surface, appear relatively unimportant, and indeed dangerous, since some of the efficacy may be lost. Once a successful therapy has been introduced, and perhaps shown to be effective in a randomised trial, it often survives largely unaltered for many years, since clinicians fear that tampering may abrogate the effect. The administration of long term maintenance therapy in acute lymphoblastic leukaemia based on a trial reported in 1963 (Freireich *et al.*, 1963) is one example among many. This conservatism, particularly regarding treatment duration, should be resisted. In the laboratory, long-term low dose therapy is a classical method for developing resistant cell lines. It also seems likely that the effectiveness of therapy given at relapse will be impaired by longer initial durations of treatment. Furthermore, the additional toxicity produced by more therapy may not merely be undesirable, it may also have implications for the patients' psychological outlook, their immune response, and thus possibly their chances of relapse (Greer & Watson, 1987).

3.2.2. Application to choice of treatment for individuals

There are two distinct approaches to application of the above principles to the choice of treatment for individuals. One is to use growth rate related prognostic factors to delineate ever smaller groups of patients with different growth rates. For instance, by choosing patients based on grade of tumour or S-phase fraction measurements in breast cancer, it is possible to target a group of early stage high risk patients with aggressive tumours, for whom intensive therapy may be appropriate (O'Reilly & Richards, 1992). Of course, within these subgroups, there will still be a wide range of different growth rates (see for example, the distribution of Ki67 values for different histological subtypes of NHL and breast cancer (Gerdes *et al.*, 1984; Gatter *et al.*, 1986; Barnard *et al.*, 1987)). If possible, a better approach would be to estimate the growth rates of individual tumours before treatment, choose dose and duration appropriately, and, ideally, monitor response, to determine when treatment was no longer effective or necessary.

Diseases where choice of treatment based on proliferation values of individual tumours might be especially appropriate and effective, are those with a wide range of growth rates, and an average growth rate which is relatively high. One such malignancy is non-Hodgkin's lymphoma. Prospective studies are unfortunately lacking, but analysis of response and duration of response in patients studied with the Ki67 antibody and labelling-index techniques suggests that, if treated early and intensively, response rates might be higher in the more rapidly proliferating tumours, and durations of response might be different (Hall *et al.*, 1987). Relapses in rapidly

growing tumours occur early, and those surviving this phase tend to be cured (or remain disease-free much longer). Relapses in slowly growing tumours may occur late, and be spread over a long time interval.

Various models have been proposed to monitor tumour response, by repeated tumour volume estimates (Birkhead & Gregory, 1984; Gregory *et al.*, 1988) or tumour marker levels (Price *et al.*, 1990a; 1990b) and to infer resistance and growth rate parameters during treatment. One such model will be described in detail in chapter 7. It is possible that such models could be used to determine when to stop treatment for an individual, following elimination of all but resistant disease, and possibly switch to an (hopefully non-cross-resistant) alternative (Gregory *et al.*, 1988; 1990).

3.3 New approaches to estimating resistance and growth rates.

Although the principles outlined above have been useful in designing treatment regimes and strategies, their application has often proved slow and laborious. It may take many trials to establish the optimum number of courses of treatment to administer, as demonstrated by, for instance, the trials already mentioned in Wilms tumour (D'Angio *et al.*, 1976; 1981; 1989), or to establish what doses are required to achieve optimum cell-kill, quite apart from the problem of which drugs to combine in the first place. This is partly because clinical trials are designed merely to discover whether one treatment is better than another, and not why it is better, and to what degree (see chapter 1). The models described in this thesis have been

designed to provide hypotheses to explain trial results in terms of some of the fundamental biological processes involved. It is hoped that by so doing, they may add to the understanding of these processes, and how therapy affects them, and so lead to more rapid development in this area. These mathematical models will now be described, and applications given.

Chapter 4

THE REMISSION DURATION MODEL

Chapter 4

THE REMISSION DURATION MODEL

4.1 Introduction

For many cancers, treatment will eradicate some or all of the measurable tumour. An apparent disease-free period ensues, often followed by relapse. There is at present, however, no way of determining how close the patient came to being cured: were there just a few cancer cells remaining when the treatment finished, or were there many cancer cells left? The length of the disease-free period does not in itself determine this, since tumours are known to grow at different rates. A late relapse could equally well be a result of rapid re-growth following near extinction of the tumour, or slow re-growth of a sizable (but not clinically detectable) residual tumour.

It is thought that treatment can influence both the amount of tumour killed and the rapidity with which this occurs, and that where treatment fails to cure the patient, this is usually because the tumour has become resistant. These ideas, originally explored by Skipper *et al* (1964), provide possible explanations for observed differences between treatments, and often form part of the rationale for design of new trials. However, as discussed in the previous chapter, there is currently no way of quantifying these effects, with the result that many trials are designed using a mixture of experience and guesswork. If it were possible to explain why one

treatment proved better than another then future trials could be designed more rationally. For example, treatment could be targeted to groups more likely to benefit from intensification, or reduced in groups seen to need little further treatment. The hypotheses raised to explain why one treatment proved better, might suggest that some choices for the next trial would be unlikely to prove beneficial, while others would be worth pursuing. Trials could be expected to lead on logically one from another, and progress might be more rapid.

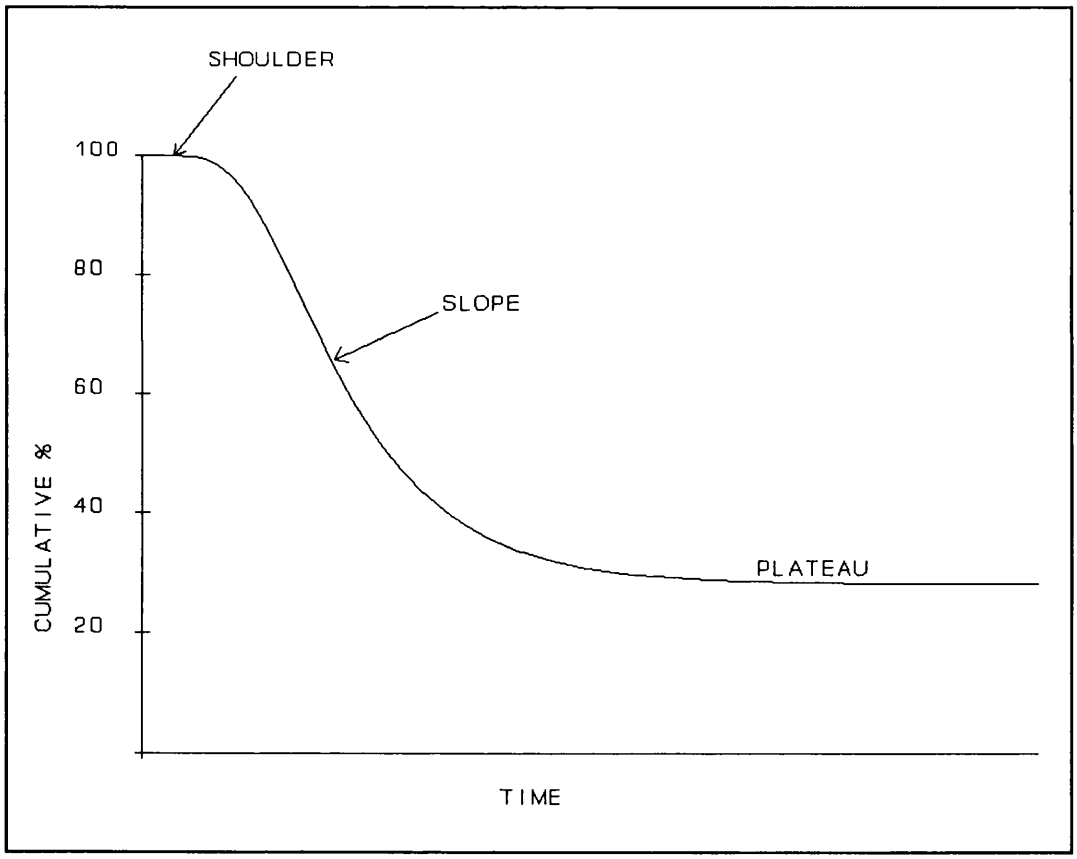


Figure 4.1 Idealised response duration curve, showing a shoulder, followed by a slope as relapses occur, and ending with a plateau

The mathematical model which has been developed seeks to derive this information from the durations of response to treatment for a group of patients. The actuarial

methods described in chapter 2 are used in the analysis of many clinical trials and treatments to plot curves showing the proportion of patients still in response over time. It was observed that the shapes of such response duration curves exhibited certain patterns. Firstly it was observed that such curves exhibited plateaus, i.e. the curves seemed to be asymptotically approaching some final percentage greater than zero. Secondly the curves seemed to have a steep slope before flattening out to the plateau, and the approximate period over which they flattened out consistently began at the same time post treatment for a particular cancer. Thirdly, response duration curves with a higher plateau also showed more of an initial shoulder. Fourthly, the remission duration curves for tumours which were thought to be faster growing had steeper slopes than those for which the tumour was thought to grow slowly. These concepts are shown diagrammatically in figure 4.1.

I thought that it might be possible to explain these observations in terms of the distribution and re-growth of the residual tumour post treatment. The curves would approach a plateau because if there was a limited range of re-growth rates, the smallest residual tumour would determine the longest relapse time. If, for a particular cancer, the ranges of growth rates were similar this would also explain why the plateaus began at approximately the same time in all the different trials and treatments for the particular cancer under study. The steep slopes on the curve would result from a clustering of volumes of residual tumour about the mean or median of the distribution of residual tumour volumes for the patient population. If the distribution of residual tumour was sufficiently constricted, a shoulder would appear on the curve when every patient in the population had their residual tumour

volume reduced well below the level where it could be detected clinically. This would be true since even for the patient with the greatest residual volume it would take some time for the tumour to re-grow to the point where it could be detected clinically, and thus a shoulder would be evident. To have caused this degree of cell-kill for the patient with the *largest* residual volume, the treatment would be more likely to cure the most susceptible patients, thus producing a higher plateau. Finally, a rapid re-growth would shorten the part of the curve with the steep slope, and thus make the slope still steeper. A mathematical model was therefore developed to try to quantify these ideas.

4.2 Basic description of the model

When evaluating a patient's response to therapy, clinicians often try to measure any reduction in volume of disease following treatment. A reduction of $>50\%$ in volume (or the product of two tumour diameters) is conventionally referred to as a partial response (Hayward *et al.*, 1978). When the tumour is reduced to a volume which can no longer be detected clinically, a 'complete' response (or remission) is documented. Other definitions are sometimes necessary, when perhaps the parameter or marker which is being used to assess the tumour, would be expected to be present in small quantities anyway. For instance, in acute myelogenous leukaemia, complete remission is defined as occurring when the bone marrow contains less than 5% of 'blast' cells (Birkhead *et al.*, 1987).

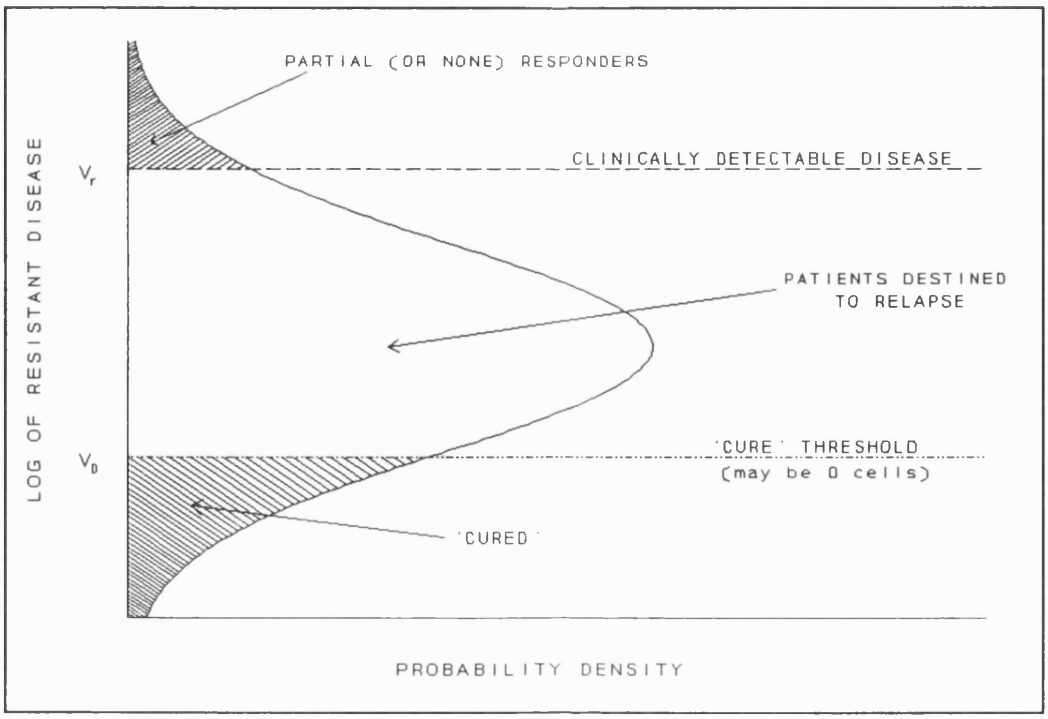


Figure 4.2 Assumed distribution of resistant disease at the start of treatment for the whole patient population.

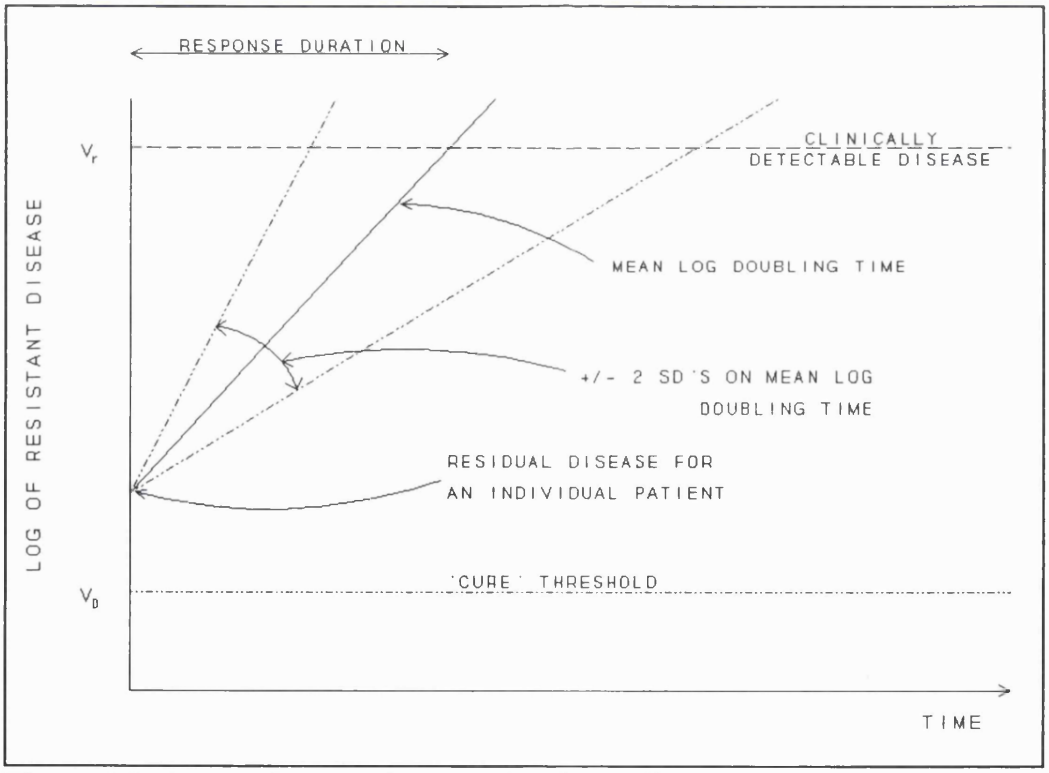


Figure 4.3 Assumed regrowth rates of resistant disease.

The first model assumption is that there exists some volume of disease V_r, below which the tumour cannot be detected clinically (or distinguished from normal tissue) for the disease being examined. (The quantification and variability in V_r is discussed later). It is further assumed that a 'complete response' occurs when the tumour volume is reduced below V_r, and that clinical relapse (re-emergence of the disease after a period of apparent disease absence) occurs when the tumour grows back to this volume or greater. It is further assumed that some volume of disease is resistant to treatment at the initiation of treatment, and that this volume of disease is log-normally distributed over the population of patients under consideration. The reasons for this assumption will be given in detail shortly. The mean and standard deviation of the log of this volume of resistant tumour will be denoted by μ_v and σ_v respectively (figure 4.2). In the event of the volume of resistant tumour being less than a given log volume, V_o, the patient is assumed to be cured (it is of course possible to assume V_o = 1 cell, ie that all tumour cells need to be eliminated for the patient to be cured). Otherwise the resistant tumour is assumed to grow exponentially during and after treatment until relapse occurs. The rate of this re-growth is assumed to be taken from a log-normal distribution of doubling times, the mean and standard deviation of the log of the doubling times being denoted by μ_g and σ_g respectively (figure 4.3). The reasons for assuming that these two distributions, of resistant tumour volumes and of doubling times, are log normal will now be discussed.

4.3 Description and justification of model assumptions

4.3.1 Tumour growth

Tumour growth is assumed to be exponential. For tumours in remission this is likely to be a good approximation (Laird, 1964; Sullivan & Salmon, 1972). The exponential assumption may be less accurate for large tumours when a Gompertzian rule may be more appropriate (see chapter 3). Thus care should be taken when applying the model to survival data, since the end-point of 10^{12} cells (a likely estimate for a 'fatal' tumour volume) is thought to involve Gompertzian kinetics (Norton & Simon, 1977). For durations of response, where relapse is often considered to result from tumour volumes of approximately 10^9 cells, an exponential assumption is likely to be adequate.

Turning to the *distribution* of the growth parameter, a wide variety of tumours reported by Shackney *et al* (1978) and Pearlman (1983) showed a log-normal distribution of doubling times (for exponential tumour growth, the time taken for the tumour to double in volume is constant, and so this measure can be used to describe the growth rate). The exponential parameter (say α) is related to the doubling time (DT) by the simple formula:

$$\alpha = \frac{\log_e 2}{DT}$$

and where doubling times are reported, exponential growth is assumed. Shackney reported measurements of growth rates showing log-normal distributions of doubling times within particular cancers. This finding held for a wide range of cancers

including breast cancer, lymphomas, adenocarcinomas of the lung and colon, and testicular carcinoma. Norton (1988) developed a mathematical model to show how different sized tumours with this growth pattern would produce systematically different response duration curves. His assumption of a log-normal distribution of doubling times (Norton, 1988), was based on analysis of breast cancer data from an untreated series of patients (Bloom *et al.*, 1962) and from patients detected after breast screening (Heuser *et al.*, 1979).

4.3.2 Composition of the residual tumour population and treatment duration

The model assumptions described to this point could apply equally well to residual or resistant tumour. In a mathematical sense, the model assumptions about volume of disease after treatment are open-ended as to what constitutes this disease. However, the assumption of log-normality for the volume of residual/resistant tumour is based on work suggesting that resistance arises as a result of mutations, which occur in a stochastic fashion at a constant rate (Goldie & Coldman, 1979). This being the case, it should apply to the resistant component of the residual tumour volume only, and not to the residual tumour volume in total.

The *duration* of treatment can therefore be important when interpreting both the model fits and differences between parameters (to be described shortly). Interpretation should be straightforward in cases where sufficient treatment has been given to eradicate all sensitive disease. It is of course difficult to be certain that this has happened, though in the main, treatments are probably overlong in duration,

since clinicians tend not to stop treatment early in case further benefit is still possible (see previous chapter). Indeed, sometimes treatment is given continuously until relapse. Where short durations of treatment have been given, or the ideal duration of treatment is problematic, interpretation may be difficult.

4.3.3 Distribution of resistant tumour volume prior to treatment

The distribution of volumes of tumour resistant to treatment is difficult to determine. However Goldie and Coldman (1979) examined the consequences of random spontaneous mutations to resistance occurring during the tumour's lifetime. This was suggested as a likely process for the development of tumour heterogeneity, and thus resistance to therapy. It will be assumed from this point onwards that resistant tumour is composed of resistant cells. Goldie and Coldman simulated the growth of a large number of tumours whose cells had constant probabilities of birth, death, and mutation, and plotted the distribution of resistant cells when the tumour had reached a given size (they in fact chose 3.2×10^4 cells). My examination (for details see below) of the Goldie and Coldman curve suggested that this distribution could be reasonably approximated by a log-normal distribution. I have thus repeated the simulation, using various different values for the birth, death and mutation parameters, to confirm this log-normality.

The simulation starts with a given number of sensitive and resistant cells, and uses random numbers to decide, for each generation of growth, whether each cell undergoes birth (with or without mutation), death or neither birth nor death(i.e. the

cell remains dormant). That is to say, all cells are assumed to act independently in such a way that each has probabilities b.Δt and d.Δt of dividing into 2 cells (birth) or being lost to the population (differentiation/death) in any small time interval Δt. Given a "birth", the new cell is assumed to have probability α.b of mutating to resistance. b and d are assumed time-independent constants with b > d (a "birth" advantage).

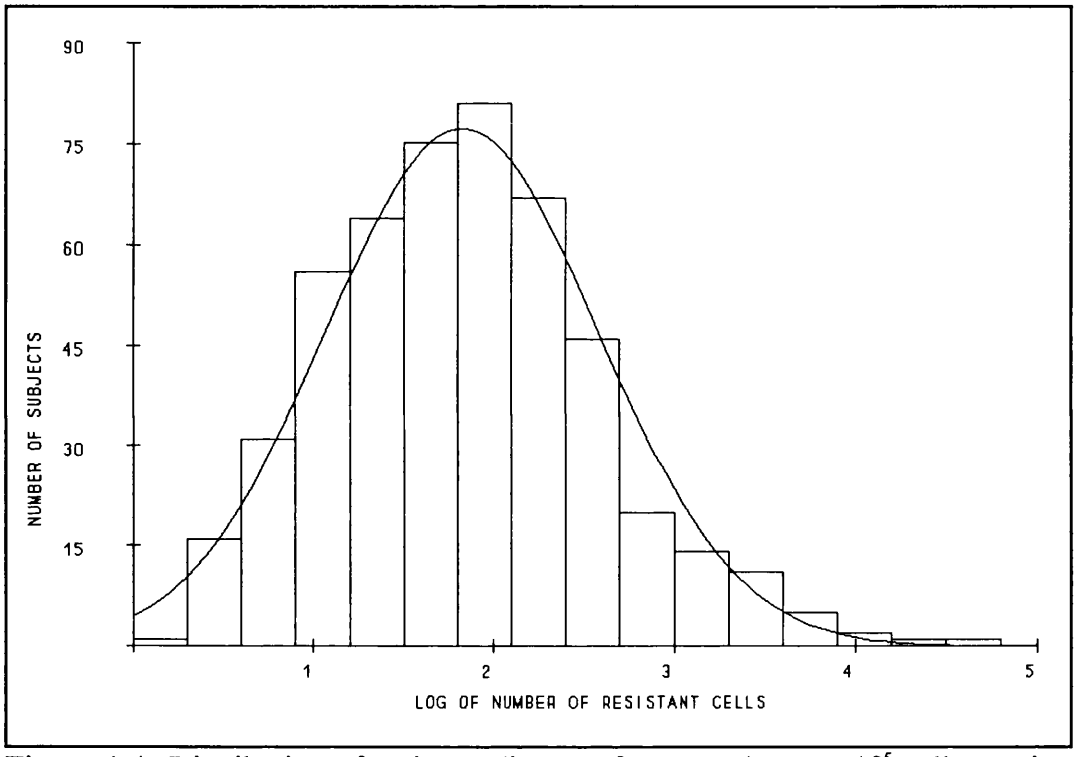


Figure 4.4 Distribution of resistant disease after growth up to 10^5 cells starting from a single sensitive cell, with parameters: $\alpha = 10^{-5}$, b=.505, d=.495.

These growth assumptions are of course somewhat artificial since they smooth out the probability of division over time (the cell cycle in fact has periods where cell division may take place, and periods where this will not happen (Rubinow & Lebowitz, 1976)). However, for the growth of large numbers of cells (involving large numbers of divisions), this makes little difference to the patterns of growth,

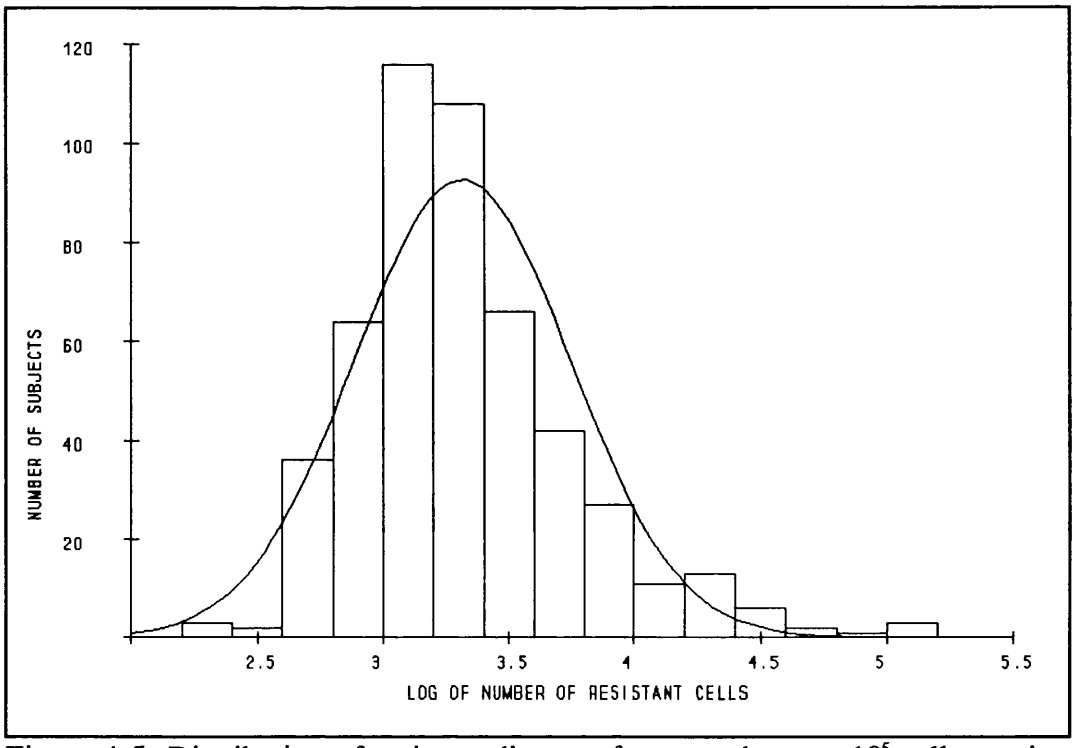


Figure 4.5 Distribution of resistant disease after growth up to 10^5 cells starting from a single sensitive cell, with parameters: $\alpha = 10^{-4}$, b=.505, d=.495.

and, for well-established populations, growth becomes approximately exponential under these rules (this was verified by examining population sizes at different points throughout the simulations). There is assumed to be no back mutation, since any effects were demonstrated to be negligible (see Goldie and Coldman (1979)). The progeny of resistant cells are therefore assumed to be also resistant. The distributions of resistant cells were plotted when the tumour had reached various sizes, from 10^5 cells upwards. The starting numbers of sensitive and resistant cells were taken to be 1 and 0 respectively. From these starting values many of the simulated tumours will die out (the probability of extinction is in fact $(d/b)^n$ for a population of n cells), thus only those that reached the specified final tumour size were examined for the numbers of resistant cells present.

The results for a variety of different values of α, b and d are shown in table 4.1. Histograms showing the distribution of the numbers of resistant cells for the set of simulated tumour growths can be produced, as shown in figure 4.4. These can be tested for normality using the Shapiro-Francia test, which correlates the values themselves with their normal scores, and examines the resulting r^2 value (Altman, 1991). Values close to 1 indicate normality. The Shapiro-Francia statistics and p-values are also given in table 4.1. Although these are, in the main, significant, the deviation from log-normality is slight, as can be seen in figure 4.5, which shows one of the worst cases. Furthermore, with the lower mutation rates (10^{-6} and less) the distributions are very close to normal, with non-significant or barely significant Shapiro-Francia statistics. Mutation rates of this magnitude and lower probably correspond to the clinical setting (Goldie & Coldman, 1979), but are difficult to test because of limitations in the speed of calculation on the available computers. The higher mutation rates (10^{-5} and greater) would lead to very large resistant volumes on presentation, and does not therefore match the cases and tumours used in the applications.

4.3.4 Starting time for measurement of response duration

If it is assumed, in view of all the points just mentioned, that the model is estimating resistant tumour volume just prior to treatment, said resistance having been acquired during the tumour's growth prior to presentation, then the response durations should be measured from this time. However, response durations are often reported from the time of response.

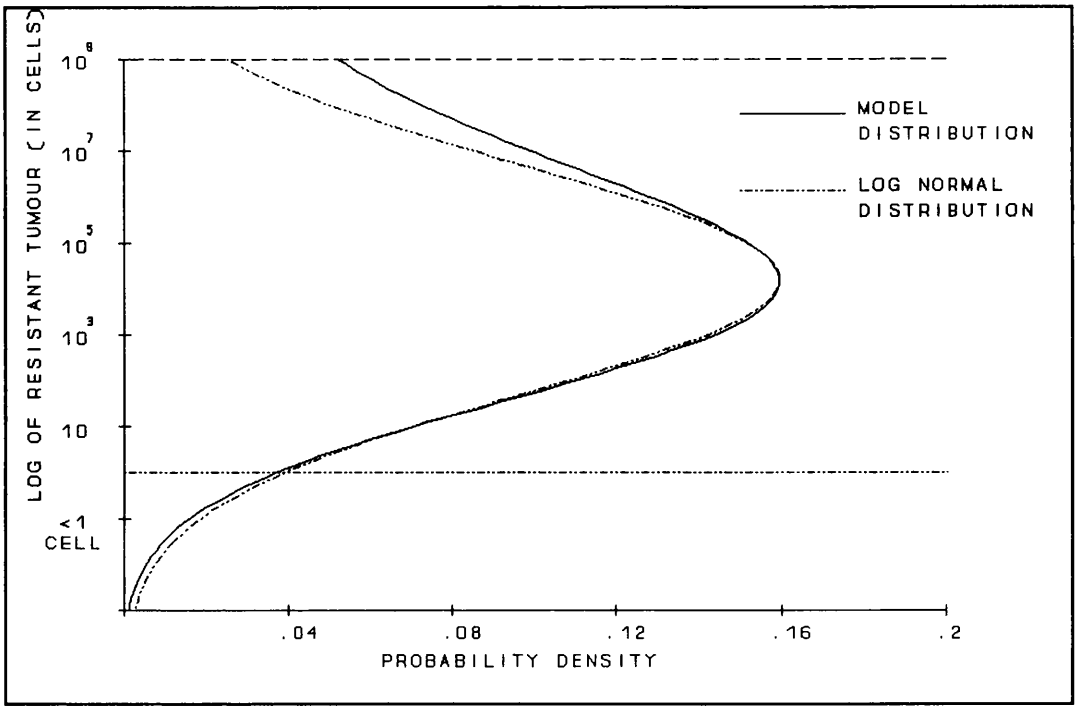


Figure 4.6 The expected distribution of resistant disease for MRC AML trial 8 (see chapter 5) after 60 days. See text for assumptions.

Fortunately, times from treatment to response are usually sufficiently short that a distribution of resistant tumour volume which was log normal at the start of treatment, will still be approximately log-normal at the time of response, if the distribution of doubling times is log normal. This is demonstrated in figure 4.6 which shows the expected distribution of the volume of resistant tumour for the first example, presented in chapter 5, after 60 days, assuming an initially log normal distribution, and that re-growth is taken from a log normal distribution of doubling times. (The parameter values for this plot are taken from table 5.1 (see next chapter), and a log normal distribution is also shown for comparison). Thus, although the model should ideally use times from the start of treatment, using times measured from the achievement of response is likely to provide a reasonable

approximation. Wherever possible, of course, times have been measured from the start of treatment, rather than from the time of response.

4.3.5 Independence of parameters

If the resistant component of a tumour is caused by mutations to resistance, then the volume of this resistance is determined by the mutation rate (which is *per generation*), and a random factor depending upon when the first resistant mutant arose. Thus the volume of resistance at presentation is related to the number of generations to presentation, which is independent of the doubling time. The model thus assumes that the regrowth rate is independent of the volume of resistant tumour.

4.3.6 Stem cells

It may appear that the assumptions of the model take no account of clonogenic or stem cells. However, the model merely assumes that the whole tumour is growing exponentially. This leaves room for hypotheses about the growth of clonogenic cells. For example, suppose that 1 in 1000 cells were really clonogenic, and causing repopulation of the tumour. Mackillop *et al* (1986) demonstrated that the growth of such a clonogenic compartment would parallel the growth of the whole tumour when the relative proportions (1:1000) were equal. Otherwise, the system would try to return to this balance, i.e. if the treatment had a greater cell-killing effect on the clonogenic population, this population would re-grow faster than the whole population until equilibrium was re-established (at 1:1000 cells). The model would

thus be reflecting growth of the clonogenic compartment, with perhaps 10^6 cells being required for relapse, rather than 10^9. The growth parameter could then be re-estimated under this assumption. Thus it is possible to apply the model under the assumption that the resistant tumour volume is composed entirely of clonogenic cells.

4.3.7 Variability in tumour volume on relapse

There is inevitably some variability in volumes of disease at relapse, since patients may wait different lengths of time before reporting their symptoms, or may notice new lumps, for instance, at different sizes. The model, as outlined, assumes that patients relapse at a particular volume of disease (V_r). However, the variability in volumes of disease at relapse is likely to be very small compared with the total re-growth time of the disease. For example, to progress from 1 to 10^9 cells takes 30 doublings, whereas 1 doubling from say $2cm^3$ to $4cm^3$ is unlikely to escape the patient's attention.

There is no reason to suppose that patients with a smaller volume of resistant tumour following initial treatment would present with larger or smaller volumes of disease on relapse. Therefore, the variability in volume of tumour at relapse seems likely to be random for a series of patients. Thus, even if there were significant differences between individual patients, this would not introduce bias into the model's estimates.

There may be some slight correlation of relapse volume with growth rate in slowly growing tumours, since it is possible that patients are more likely to detect rapid changes in size of lumps than slow increases. However, even in the most likely candidate for such a correlation, namely breast cancer, when such a correlation was looked for, only a very slight correlation ($r=0.29$) was observed (Brown *et al.*, 1987). It is worth noting that an r-value of 0.29 corresponds to an r^2 of .08, and thus less than 10% of the variability in relapse volumes can be attributable to differences in tumour growth rate. For all these reasons, the assumption of a constant tumour volume on relapse seems adequate.

4.3.8 Tumour doubling times

If the model is to be used to estimate tumour doubling times an assumption must be made about the volume of resistant tumour below which a "cure" is achieved. It is simplest to assume that this is one cell, i.e. that all tumour cells need to be eliminated in order to cure the patient. However, it may be that small tumours can die out naturally, or that host defence mechanisms can destroy some tumour cells. An alternative approach, therefore, is to assume a range of values of doubling times from published estimates. By using these values to fix the mean log of the doubling time, the model can be used to assess the "cure" threshold. Both these methods are considered in the examples to be given in the next chapter.

Before going on to the mathematical description of the model, it is worth repeating that the model assumptions are:

1). That there is a log-normal distribution of resistant disease after treatment, and that for a given individual having any such resistant disease remaining (some patients may have all their disease eliminated by the treatment), subsequent growth of this disease occurs until it can be detected clinically, and relapse is documented.

2) That (exponential) growth of resistant disease is taken from a log-normal distribution of doubling times.

3) That resistant disease and growth rates are independent.

The mathematics necessary to estimate the parameters of these two log-normal distributions, and thus to fit the model to real response duration times, will now be described.

4.4 Mathematical description of the model

4.4.1 Likelihood derivation

Consider the following representation of the assumptions just described.

Let the random variables V and G be normally distributed, and represent the log of the resistant tumour volume (composed of resistant cells), with mean μ_v and standard

deviation σ_v, and the log of the tumour doubling time, with mean μ_g and standard deviation σ_g, i.e. $V \sim N(\mu_v,\sigma_v)$ and $G \sim N(\mu_g,\sigma_g)$. Consider an individual tumour,i, with a log resistant volume v below the log relapse threshold, V_r, and having a log tumour doubling time g. For notational simplicity let N_g denote the normal distribution function value at g when the mean is μ_g and the standard deviation is σ_g, i.e.

$$N_g = (1/\sigma_g\sqrt{2\pi}) \exp(-(g-\mu_g)^2/2\sigma_g^2)$$

Then the probability, P_g say, of relapse before a given time t for this patient, is given by integrating over all values of g which result in relapse before t, i.e.

$$P_g = \int_{-\infty}^{\log_e(t/(V_r-v))+\log_e(\log_e(2))} N_g \, dg \qquad \{4.1\}$$

The upper limit of integration in the preceding equation is found by considering the growth of the log resistant tumour volume v, with a doubling time (say DT) such that relapse occurs at time t. Thus

$$\alpha t = V_r - v,$$

where α is the exponential growth parameter; α is thus related to the doubling time by the following equation:

$$\alpha = \log_e(2)/DT$$

and therefore $DT = \{t/(V_r-v)\}\log_e(2)$ and thus

$$\log_e(DT) = g = \log_e(t/(V_r-v))+\log_e(\log_e(2)) \qquad \{4.2\}$$

Having thus described the probability of relapse for a patient with some particular resistant volume v, we can extend the analysis to the whole population by noting that

the resistant tumour volume is assumed to be log-normally distributed. With this log-normality assumption, some patients may have resistant tumours whose volumes are not below the relapse threshold, V_r, as shown by the upper shaded section of figure 4.2. The proportion of those patients which are, usually termed the remitters or responders, will be denoted P(CR). Again let N_v denote the normal distribution function value at v when the mean is μ_v and the standard deviation is σ_v, i.e.

$$N_v = (1/\sigma_v\sqrt{2\pi}) \exp(-(v-\mu_v)^2/2\sigma_v^2)$$

Then the probability, P of relapse before a given time t for the whole population (of remitters) is

$$P = \frac{\int_{V_o}^{V_r} N_v \int_{-\infty}^{\log_e(t/(V_r-v))+\log_e(\log_e(2))} N_g \, dg \, dv}{P(CR)} \qquad \{4.3\}$$

$$= P(t,\mu_v,\sigma_v,\mu_g,\sigma_g), \text{ say.}$$

The integral of a normal distribution cannot be evaluated analytically. However, numerical integration methods can be applied (see section 4.6), and thus it is possible to proceed with the normal distributions suggested on both theoretical and experimental grounds, as discussed in section 4.3.

P(CR) is the probability that the resistant tumour volume was lower than the relapse threshold V_r, i.e.

$$P(CR) = \int_{-\infty}^{V_r} N_v \, dv \qquad \{4.4\}$$

This normal distribution integral can also be evaluated by numerical methods.

The probability density function (pdf) $p(t,\mu_g,\sigma_g,\mu_v,\sigma_v) = \frac{d}{dt}P(t,\mu_g,\sigma_g,\mu_v,\sigma_v)$ can be derived from equation {4.3} by differentiating under the double integral sign. The terms N_v and P(CR) do not involve t, and so it is merely necessary to differentiate P_g with respect to t. From equation {4.2}

$$g = \log_e(t) - \log_e(V_r-v) + \log_e(\log_e(2))$$

Hence

$$dg = \frac{dt}{t}$$

and thus

$$P_g = \int_{-\infty}^{\log_e(t/(V_r-v))+\log_e(\log_e(2))} N_g \, dg \qquad = \int_{0}^{t} N_g \, (1/t) \, dt$$

whence

$$\frac{dP_g}{dt} = N_g\,(1/t)$$

and therefore

$$\frac{dP}{dt} = P' = \frac{\int_{V_o}^{V_r} N_v N_g\,(1/t)\,dv}{P(CR)} \qquad \{4.5\}$$

(Note $N_g = (1/\sigma_g\sqrt{2\pi})\exp(-(g-\mu_g)^2/2\sigma_g^2)$

$= (1/\sigma_g\sqrt{2\pi})\exp(-(\log_e(t/(V_r-v))+\log_e(\log_e(2))-\mu_g)^2/2\sigma_g^2)$

and N_g is therefore a function of t, μ_g, σ_g and v).

Let the upper limit of integration in equation {4.3} be U_t, i.e.

$U_t = \log_e(t/(V_r-v))+\log_e(\log_e(2))$. Then P can be written as

$$P = \frac{\int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\,dg\,dv}{P(CR)} \qquad \{4.6\}$$

Having derived the pdf, the likelihood, L, of the data under this distribution can be evaluated. The likelihood is the product of the pdf probabilities for each completed time to relapse, and the probability of remaining in remission longer than the time under consideration for the censored times (which includes the probability of being cured, as well as the probability of relapse occurring after this time), i.e.

$$L(\mu_v,\sigma_v,\mu_g,\sigma_g) = \prod_{i=1}^{m} p(t_i,\mu_g,\sigma_g,\mu_v,\sigma_v) \prod_{i=m+1}^{n} (1-P(t_i,\mu_g,\sigma_g,\mu_v,\sigma_v))$$

where t_i, $i=1,\ldots,m$ are the completed times to relapse, and t_i, $i=m+1,\ldots,n$ are the censored times to relapse.

The likelihood can be used to estimate the parameters of the model, by altering the parameters so as to maximise the likelihood. Being a product of many individual pdfs, the likelihood of a large data set is virtually certain to be very small, hence for computational reasons, it is easier to work with the log likelihood. This has the additional advantage of being more analytically tractable. Thus

$$\text{Log } L = \sum_{i=1}^{m} \log(p(t_i,\mu_g,\sigma_g,\mu_v,\sigma_v)) + \sum_{i=m+1}^{n} \log(1-P(t_i,\mu_g,\sigma_g,\mu_v,\sigma_v)) \qquad \{4.7\}$$

To maximise a function, it is necessary to show that the first derivative is zero, that the second derivative is not zero (ie it is not a point of inflection) and that no other points obeying the first two rules have a greater value. For a function of 3 variables, this is analogous to looking for the highest mountain peak in a given terrain. If a random starting point is taken, the derivatives can be used to climb to the top of the nearest peak using Newton's method (Beale, 1988) or variants thereof.

Second derivatives make the convergence routines much more rapid, though convergence using semi-Newton methods can be obtained using first derivatives only (Beale, 1988). These two possibilities are compared in section 4.6.

4.4.2 First partial derivatives

It is therefore necessary to derive the first partial derivatives,

$$\frac{\partial \text{Log } L}{\partial \mu_v}, \quad \frac{\partial \text{Log } L}{\partial \sigma_v}, \quad \frac{\partial \text{Log } L}{\partial \mu_g}, \quad \frac{\partial \text{Log } L}{\partial \sigma_g},$$

From {4.7} it is clear that the complete and incomplete times can be taken separately. Then

$$\frac{\partial}{\partial \theta_i} \log(1-P) = -\frac{1}{(1-P)} \frac{\partial P}{\partial \theta_i} \quad \text{and} \quad \frac{\partial}{\partial \theta_i} \log(P') = \frac{1}{P'} \frac{\partial P'}{\partial \theta_i} \quad (i=1,4)$$

where $\theta_1=\mu_v$, $\theta_2=\sigma_v$, $\theta_3=\mu_g$, $\theta_4=\sigma_g$. It is therefore necessary to differentiate P and P′ with respect to each of the four parameters $\mu_v,\sigma_v,\mu_g,\sigma_g$. Starting with μ_v and differentiating under the integral given in equation {4.6}

$$\frac{\partial P}{\partial \mu_v} = \frac{\int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\} N_v \int_{-\infty}^{U_t} N_g \, dg \, dv}{P(CR)}$$

$$- \frac{\frac{\partial (P(CR))}{\partial \mu_v} \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g \, dg \, dv}{(P(CR))^2} \qquad \{4.8\}$$

where, from equation {4.4} differentiating under the integral,

$$\frac{\partial (P(CR))}{\partial \mu_v} = \int_{-\infty}^{V_r} \{(v-\mu_v)/\sigma_v^2\} N_v \, dv \qquad \{4.9\}$$

Similarly, from equation {4.5} differentiating under the integral

$$\frac{\partial P'}{\partial \mu_v} = \frac{\int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\}\, N_v\, N_g\, (1/t)\, dv}{P(CR)}$$

$$- \frac{\dfrac{\partial (P(CR))}{\partial \mu_v} \int_{V_o}^{V_r} N_v\, N_g\, (1/t)\, dv}{(P(CR))^2} \qquad \{4.10\}$$

Similarly differentiating P with respect to σ_v from equation {4.6}

$$\frac{\partial P}{\partial \sigma_v} = \frac{\int_{V_o}^{V_r} [\{(v-\mu_v)^2/\sigma_v^3\}-1/\sigma_v]\, N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{P(CR)}$$

$$- \frac{\dfrac{\partial (P(CR))}{\partial \sigma_v} \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2} \qquad \{4.11\}$$

where, from equation {4.4} differentiating under the integral,

$$\frac{\partial (P(CR))}{\partial \sigma_v} = \int_{-\infty}^{V_r} [\{(v-\mu_v)^2/\sigma_v^3\}-1/\sigma_v]\, N_v\, dv \qquad \{4.12\}$$

Differentiating P′ with respect to σ_v from equation {4.5}

$$\frac{\partial P'}{\partial \sigma_v} = \frac{\int_{V_o}^{V_r} [\{(v-\mu_v)^2/\sigma_v^3\}-1/\sigma_v]\, N_v\, N_g\, (1/t)\, dv}{P(CR)}$$

$$- \frac{\frac{\partial (P(CR))}{\partial \sigma_v} \int_{V_o}^{V_r} N_v\, N_g\, (1/t)\, dv}{(P(CR))^2} \qquad \{4.13\}$$

Differentiating P with respect to μ_g from equation {4.6}

$$\frac{\partial P}{\partial \mu_g} = \frac{\int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} \{(g-\mu_g)/\sigma_g^2\}\, N_g\, dg\, dv}{P(CR)}$$

$$- \frac{\frac{\partial (P(CR))}{\partial \mu_g} \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2}$$

which reduces to

$$- \frac{\int_{V_o}^{V_r} N_v\, N_g\, dv}{P(CR)} \qquad \{4.14\}$$

since $\frac{\partial (P(CR))}{\partial \mu_g}$ is zero.

Differentiating P′ with respect to μ_g from equation {4.5}, noting that $\frac{\partial(P(CR))}{\partial \mu_g}$ is zero gives

$$\frac{\partial P'}{\partial \mu_g} = \frac{\int_{V_o}^{V_r} \{(g-\mu_g)/\sigma_g^2\}\, N_v\, N_g\, (1/t)\, dv}{P(CR)} \qquad \{4.15\}$$

For the last of the four first derivatives, differentiating P with respect to σ_g from equation {4.6} gives

$$\frac{\partial P}{\partial \sigma_g} = \frac{\int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} (1/\sigma_g)[\{(g-\mu_g)/\sigma_g\}^2-1]\, N_g\, dg\, dv}{P(CR)}$$

$$- \frac{\frac{\partial(P(CR))}{\partial \sigma_g} \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2}$$

Since $\frac{\partial(P(CR))}{\partial \sigma_g}$ is zero, the second term is eliminated, hence

$$\frac{\partial P}{\partial \sigma_g} = \frac{\int_{V_o}^{V_r} N_v\, (1/\sigma_g) \int_{-\infty}^{U_t} [\{(g-\mu_g)/\sigma_g\}^2-1]\, N_g\, dg\, dv}{P(CR)}$$

which reduces to

$$- \frac{\int_{V_o}^{V_r} \{(g-\mu_g)/\sigma_g\}\, N_v\, N_g\, dv}{P(CR)} \qquad \{4.16\}$$

Finally, differentiating P′ with respect to σ_g from equation {4.5}, again noting that $\frac{\partial(P(CR))}{\partial\sigma_g}$ is zero gives

$$\frac{\partial P'}{\partial \sigma_g} = \frac{\int_{V_o}^{V_r} (1/\sigma_g)[\{(g-\mu_g)/\sigma_g\}^2-1]\, N_v\, N_g\, (1/t)\, dv}{P(CR)} \qquad \{4.17\}$$

4.4.3 Second partial derivatives

For the second partial derivatives note initially that

$$\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}\log(1-P) = -\frac{1}{(1-P)}\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}(P) - \frac{1}{(1-P)^2}\left(\frac{\partial P}{\partial\theta_i}\frac{\partial P}{\partial\theta_j}\right) \quad \text{and}$$

$$\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}\log(P') = \frac{1}{P'}\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}(P') - \frac{1}{P'^2}\left(\frac{\partial P'}{\partial\theta_i}\frac{\partial P'}{\partial\theta_j}\right) \qquad (i=1,4;\ j=1,4)$$

where $\theta_1=\mu_v$, $\theta_2=\sigma_v$, $\theta_3=\mu_g$, $\theta_4=\sigma_g$, which for i=j simplifies to

$$\frac{\partial^2}{\partial\theta_i^2}\log(1-P) = -\frac{1}{(1-P)}\frac{\partial^2 P}{\partial\theta_i^2} - \frac{1}{(1-P)^2}\left(\frac{\partial P}{\partial\theta_i}\right)^2 \quad \text{and}$$

$$\frac{\partial^2}{\partial\theta_i^2}\log(P') = \frac{1}{P'}\frac{\partial^2 P'}{\partial\theta_i^2} - \frac{1}{P'^2}\left(\frac{\partial P'}{\partial\theta_i}\right)^2 \qquad (i=1,4)$$

It is therefore necessary to form the 16 second partial derivatives of P and P′ with respect to the four parameters $\mu_v,\sigma_v,\mu_g,\sigma_g$. However, because of the symmetry where $\frac{\partial P}{\partial\theta_i}\frac{\partial P}{\partial\theta_j} = \frac{\partial P}{\partial\theta_j}\frac{\partial P}{\partial\theta_i}$ it is only necessary to derive 10 second partial derivatives.

Taking the second partial derivatives in turn, and differentiating under the integrals in equations {4.8} to {4.17}

$$\frac{\partial^2 P}{\partial \mu_v^2} = \frac{\int_{V_o}^{V_r} (1/\sigma_v^2)[\{(v-\mu_v)^2/\sigma_v^2\}-1]\, N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{P(CR)}$$

$$- \frac{2\,\frac{\partial (P(CR))}{\partial \mu_v} \int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2}$$

$$- \frac{\frac{\partial^2 (P(CR))}{\partial \mu_v^2} \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2}$$

$$+ \frac{2\left(\frac{\partial (P(CR))}{\partial \mu_v}\right)^2 \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^3} \qquad \{4.18\}$$

where, from equation {4.9} differentiating under the integral,

$$\frac{\partial^2 (P(CR))}{\partial \mu_v^2} = \int_{-\infty}^{V_r} (1/\sigma_v^2)[\{(v-\mu_v)^2/\sigma_v^2\}-1]\, N_v\, dv \qquad \{4.19\}$$

Similarly for P′

$$\frac{\partial^2 P'}{\partial \mu_v^2} = \frac{\int_{V_o}^{V_r} (1/\sigma_v^2)[\{(v-\mu_v)^2/\sigma_v^2\}-1]\, N_v\, N_g\, (1/t)\, dv}{P(CR)}$$

$$- \frac{2\, \frac{\partial (P(CR))}{\partial \mu_v} \int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\}\, N_v\, N_g\, (1/t)\, dv}{(P(CR))^2}$$

$$- \frac{\frac{\partial^2 (P(CR))}{\partial \mu_v^2} \int_{V_o}^{V_r} N_v\, N_g\, (1/t)\, dv}{(P(CR))^2}$$

$$+ \frac{2\left(\frac{\partial (P(CR))}{\partial \mu_v}\right)^2 \int_{V_o}^{V_r} N_v\, N_g\, (1/t)\, dv}{(P(CR))^3} \qquad \{4.20\}$$

where $\frac{\partial^2 (P(CR))}{\partial \mu_v^2}$ is given in equation {4.19}.

Turning to σ_v

$$\frac{\partial^2 P}{\partial \sigma_v^2} = \frac{\int_{V_o}^{V_r} (1/\sigma_v^2)[\{(v-\mu_v)/\sigma_v\}^4 - \{(v-\mu_v)/\sigma_v\}^2 + 2]\, N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{P(CR)}$$

$$- \frac{2\,\frac{\partial (P(CR))}{\partial \sigma_v} \int_{V_o}^{V_r} (1/\sigma_v)[\{(v-\mu_v)/\sigma_v^2\}^2 - 1]\, N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2}$$

$$- \frac{\frac{\partial^2 (P(CR))}{\partial \sigma_v^2} \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2}$$

$$+ \frac{2\left(\frac{\partial (P(CR))}{\partial \sigma_v}\right)^2 \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^3} \qquad \{4.21\}$$

where, from equation {4.9} differentiating under the integral,

$$\frac{\partial^2 (P(CR))}{\partial \sigma_v^2} = \int_{-\infty}^{V_r} (1/\sigma_v^2)[\{(v-\mu_v)/\sigma_v\}^4 - \{(v-\mu_v)/\sigma_v\}^2 + 2]\, N_v\, dv \qquad \{4.22\}$$

$$\frac{\partial^2 P'}{\partial \sigma_v^2} = \frac{\int_{V_o}^{V_r} (1/\sigma_v^2)[\{(v-\mu_v)/\sigma_v\}^4 - \{(v-\mu_v)/\sigma_v\}^2 + 2]\, N_v\, N_g\, (1/t)\, dv}{P(CR)}$$

$$- \frac{2\, \frac{\partial (P(CR))}{\partial \sigma_v} \int_{V_o}^{V_r} (1/\sigma_v)[\{(v-\mu_v)/\sigma_v^2\}^2 - 1]\, N_v\, N_g\, (1/t)\, dv}{(P(CR))^2}$$

$$- \frac{\frac{\partial^2 (P(CR))}{\partial \sigma_v^2} \int_{V_o}^{V_r} N_v\, N_g\, (1/t)\, dv}{(P(CR))^2}$$

$$+ \frac{2\left(\frac{\partial (P(CR))}{\partial \sigma_v}\right)^2 \int_{V_o}^{V_r} N_v\, N_g\, (1/t)\, dv}{(P(CR))^3} \qquad \{4.23\}$$

where $\frac{\partial^2 (P(CR))}{\partial \sigma_v^2}$ is given in equation {4.22}.

$$\frac{\partial}{\partial\sigma_v}\frac{\partial P}{\partial\mu_v} = \frac{\int_{V_o}^{V_r}(1/\sigma_v^2)\{(v-\mu_v)/\sigma_v\}[\{(v-\mu_v)/\sigma_v\}^2-3]\,N_v \int_{-\infty}^{U_t} N_g\,dg\,dv}{P(CR)}$$

$$-\frac{\frac{\partial(P(CR))}{\partial\mu_v}\int_{V_o}^{V_r}(1/\sigma_v)[\{(v-\mu_v)/\sigma_v^2\}^2-1]\,N_v \int_{-\infty}^{U_t} N_g\,dg\,dv}{(P(CR))^2}$$

$$-\frac{\frac{\partial(P(CR))}{\partial\sigma_v}\int_{V_o}^{V_r}\{(v-\mu_v)/\sigma_v^2\}\,N_v \int_{-\infty}^{U_t} N_g\,dg\,dv}{(P(CR))^2}$$

$$-\frac{\frac{\partial}{\partial\sigma_v}\frac{\partial(P(CR))}{\partial\mu_v}\int_{V_o}^{V_r}N_v \int_{-\infty}^{U_t} N_g\,dg\,dv}{(P(CR))^2}$$

$$+\frac{2\,\frac{\partial(P(CR))}{\partial\mu_v}\,\frac{\partial(P(CR))}{\partial\sigma_v}\int_{V_o}^{V_r}N_v \int_{-\infty}^{U_t} N_g\,dg\,dv}{(P(CR))^3} \qquad \{4.24\}$$

where, from equation {4.9} differentiating under the integral,

$$\frac{\partial}{\partial\sigma_v}\frac{\partial(P(CR))}{\partial\mu_v} = \int_{-\infty}^{V_r}(1/\sigma_v^2)\{(v-\mu_v)/\sigma_v\}[\{(v-\mu_v)/\sigma_v\}^2-3]\,N_v\,dv \qquad \{4.25\}$$

$$\frac{\partial}{\partial\sigma_v}\frac{\partial P'}{\partial\mu_v} = \frac{\int_{V_o}^{V_r}(1/\sigma_v^2)\{(v-\mu_v)/\sigma_v\}[\{(v-\mu_v)/\sigma_v\}^2-3]\,N_v\,N_g\,(1/t)\,dv}{P(CR)}$$

$$-\frac{\frac{\partial(P(CR))}{\partial\mu_v}\int_{V_o}^{V_r}(1/\sigma_v)[\{(v-\mu_v)/\sigma_v^2\}^2-1]\,N_v\,N_g\,(1/t)\,dv}{(P(CR))^2}$$

$$-\frac{\frac{\partial(P(CR))}{\partial\sigma_v}\int_{V_o}^{V_r}\{(v-\mu_v)/\sigma_v^2\}\,N_v\,N_g\,(1/t)\,dv}{(P(CR))^2}$$

$$-\frac{\frac{\partial}{\partial\sigma_v}\frac{\partial(P(CR))}{\partial\mu_v}\int_{V_o}^{V_r}N_v\,N_g\,(1/t)\,dv}{(P(CR))^2}$$

$$+\frac{2\,\frac{\partial(P(CR))}{\partial\mu_v}\,\frac{\partial(P(CR))}{\partial\sigma_v}\int_{V_o}^{V_r}N_v\,N_g\,(1/t)\,dv}{(P(CR))^3} \qquad \{4.26\}$$

where $\frac{\partial}{\partial\sigma_v}\frac{\partial(P(CR))}{\partial\mu_v}$ is given in equation {4.25}.

The other second partial derivatives are simpler, since P(CR) does not vary with respect to μ_g or σ_g, and thus the extra terms related to P(CR) are all zero. Thus, differentiating under the integral with respect to μ_v from equation {4.14} gives:

$$\frac{\partial}{\partial \mu_v}\frac{\partial P}{\partial \mu_g} = -\frac{\int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\}\, N_v\, N_g\, dv}{P(CR)} \qquad \{4.27\}$$

and from equation {4.15}:

$$\frac{\partial}{\partial \mu_v}\frac{\partial P'}{\partial \mu_g} = \frac{\int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\}\, \{(g-\mu_g)/\sigma_g^2\}\, N_v\, N_g\, (1/t)\, dv}{P(CR)} \qquad \{4.28\}$$

From equation {4.16}:

$$\frac{\partial}{\partial \mu_v}\frac{\partial P}{\partial \sigma_g} = -\frac{\int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\}\, \{(g-\mu_g)/\sigma_g\}\, N_v\, N_g\, dv}{P(CR)} \qquad \{4.29\}$$

and from equation {4.17}:

$$\frac{\partial}{\partial \mu_v}\frac{\partial P'}{\partial \sigma_g} = \frac{\int_{V_o}^{V_r} \{(v-\mu_v)/\sigma_v^2\}\, (1/\sigma_g)[\{(g-\mu_g)/\sigma_g\}^2-1]\, N_v\, N_g\, (1/t)\, dv}{P(CR)} \qquad \{4.30\}$$

Similarly for σ_v differentiating under the integral from equation {4.14} gives:

$$\frac{\partial}{\partial\sigma_v}\frac{\partial P}{\partial\mu_g} = -\frac{\int_{V_o}^{V_r}[\{(v-\mu_v)/\sigma_v^3\}-1/\sigma_v]\,N_v\,N_g\,dv}{P(CR)} \qquad \{4.31\}$$

and from equation {4.15}:

$$\frac{\partial}{\partial\sigma_v}\frac{\partial P'}{\partial\mu_g} = \frac{\int_{V_o}^{V_r}[\{(v-\mu_v)/\sigma_v^3\}-1/\sigma_v]\,\{(g-\mu_g)/\sigma_g^2\}\,N_v\,N_g\,(1/t)\,dv}{P(CR)} \qquad \{4.32\}$$

From equation {4.16}:

$$\frac{\partial}{\partial\sigma_v}\frac{\partial P}{\partial\sigma_g} = -\frac{\int_{V_o}^{V_r}[\{(v-\mu_v)/\sigma_v^3\}-1/\sigma_v]\,\{(g-\mu_g)/\sigma_g\}\,N_v\,N_g\,dv}{P(CR)} \qquad \{4.33\}$$

and from equation {4.17}:

$$\frac{\partial}{\partial\sigma_v}\frac{\partial P'}{\partial\sigma_g} = \frac{\int_{V_o}^{V_r}[\{(v-\mu_v)/\sigma_v^3\}-1/\sigma_v]\,(1/\sigma_g)[\{(g-\mu_g)/\sigma_g\}^2-1]\,N_v\,N_g\,(1/t)\,dv}{P(CR)} \qquad \{4.34\}$$

Similarly for μ_g differentiating under the integral from equation {4.14} gives:

$$\frac{\partial^2 P}{\partial \mu_g^2} = - \frac{\int_{V_o}^{V_r} \{(g-\mu_g)/\sigma_g^2\} N_v N_g \, dv}{P(CR)} \qquad \{4.35\}$$

and from equation {4.15}:

$$\frac{\partial^2 P'}{\partial \mu_g^2} = \frac{\int_{V_o}^{V_r} (1/\sigma_g^2) [\{(g-\mu_g)/\sigma_g\}^2-1] N_v N_g (1/t) \, dv}{P(CR)} \qquad \{4.36\}$$

From equation {4.16}:

$$\frac{\partial^2 P}{\partial \sigma_g^2} = - \frac{\int_{V_o}^{V_r} \{(g-\mu_g)/\sigma_g^2\} [\{(g-\mu_g)/\sigma_g\}^2-2] N_v N_g \, dv}{P(CR)} \qquad \{4.37\}$$

and from equation {4.17}:

$$\frac{\partial^2 P'}{\partial \sigma_g^2} = \frac{\int_{V_o}^{V_r} (1/\sigma_g^2) [\{(g-\mu_g)/\sigma_g\}^4-5\{(g-\mu_g)/\sigma_g\}^2+2] N_v N_g (1/t) \, dv}{P(CR)} \qquad \{4.38\}$$

Finally, differentiating under the integral from equation {4.14} gives:

$$\frac{\partial}{\partial\sigma_g}\frac{\partial P}{\partial\mu_g} = \frac{\int_{V_o}^{V_r} (1/\sigma_g)[\{(g-\mu_g)/\sigma_g\}^2-1]\, N_v\, N_g\, dv}{P(CR)} \qquad \{4.39\}$$

and from equation {4.15}:

$$\frac{\partial}{\partial\sigma_g}\frac{\partial P'}{\partial\mu_g} = \frac{\int_{V_o}^{V_r} \{(g-\mu_g)/\sigma_g^3\}\, [\{(g-\mu_g)/\sigma_g\}^2-3]\, N_v\, N_g\, (1/t)\, dv}{P(CR)} \qquad \{4.40\}$$

4.4.4 Newton's method for fitting the model

Given a general likelihood function $L(\theta)$ for a column vector θ, Newton's method of obtaining the maximum likelihood is based on the first order Taylor series expansion of $G(\theta) = \partial \log L(\theta)/\partial\theta$. Given a trial value for the maximum, say θ_0, the first partial derivative vector at $\hat{\theta}$ can be written

$$G(\hat{\theta}) = G(\theta_0) - I(\theta^*)(\hat{\theta}-\theta_0)$$

where $I(\theta)$ is the Information matrix (the matrix of second partial derivatives) at θ, and where θ^* lies "between" θ_0 and $\hat{\theta}$. For θ_0 in the vicinity of $\hat{\theta}$, $I(\theta^*)$ can be approximated by $I(\theta_0)$. Thus setting $G(\hat{\theta})=0$ and solving gives

$$\hat{\theta} = \theta_0 + I(\theta_0)^{-1}G(\theta_0) \qquad \{4.41\}$$

The right side of equation {4.41} gives a new trial value for θ with which the process is repeated until successive θ estimates agree to a specified extent, and of course $G(\theta)$ should equal 0 at convergence. Note that the procedure produces, as a

by-product, the observed information matrix evaluated at the maximum likelihood estimate, $I(\hat{\theta})$ (see section 4.5.5).

Thus, given the likelihood derivation, and the first and second partial derivatives described in the previous 3 sections, Newton's method can be used to maximise the likelihood, and thus fit the model.

4.4.5 Parametric relevance of the "cure" threshold

The distribution of P has been portrayed as depending on just the four parameters μ_v, σ_v, μ_g, σ_g, and not on the volume of disease needed to be eradicated to cure the patient, V_r-V_o. This is because it is the shape of the distributions that determines the relapse pattern, and this is independent of V_r-V_o (the distribution of the resistant tumour volume can be "stretched" or "squashed" by changing V_r-V_o, but it is easily seen that a corresponding change in the tumour doubling time will produce an identical response duration curve). Some particular choices for V_o and V_r are necessary in order to fill in the whole picture and provide numerical values for the log mean doubling time, the log mean volume of resistant tumour etc. However, this is not necessary for plotting the model fit to the actuarial curve. Parameters can be estimated as a function of V_r-V_o. For instance, for the model fit given in figures 5.3 and 5.4 (see chapter 5), where V_o was assumed to be 1 cell, and V_r was assumed to be 10^9 cells, the model parameters were given as:

$\mu_v = 2.3$, $\mu_g = 1.8$, $\sigma_v = 14$, $\sigma_g = .29$

they could instead have been given as:

$\mu_v = 0.256(V_r-V_o)$, $\sigma_v = 0.2(V_r-V_o)$, $\mu_g = 126/(V_r-V_o)$, $\sigma_g = .29$

with, in this case, $(V_r-V_o)=9$.

If, for example, we were to consider that resistant tumour stem cells were the cause of relapse, and that only 10^6 tumour stem cells were required for relapse rather than 10^9 resistant tumour cells (see section 4.3.6), then the parameters (as applied to the stem cell population) would be

$V_r-V_o = 6$, and thus $\mu_v = 1.536$, $\sigma_v = 1.2$, $\mu_g = 21$, $\sigma_g = .29$

4.5 Application Methodology

The model is calibrated for a particular data set using the methodology described above. The data set simply comprises the response duration times for the group of patients of interest. A starting estimate (or guess) is made for the four model parameters, and an iterative computer procedure using Newton's method, as just described, then hopefully converges on the best fit. An example will be given shortly (see section 4.6.2).

It is difficult to be certain that the final 'maximum' found is indeed a unique maximum likelihood (and 'best fit') curve. There are several approaches to verifying the uniqueness of the fit.

Contour plots can be drawn to examine further the shape of the likelihood terrain (see section 4.5.1). A further validation is obtained by choosing widely differing starting estimates for the parameters and checking that they converge on the same

maximum. This latter method has been tried for many of the plots and estimates shown in this thesis. Only one alternative maximum was ever found, and in this case the likelihood was very much worse for one fit than for the other, and this could be easily seen by inspection of the model fit graphs. The methodology for drawing likelihood contours will now be described.

4.5.1 Likelihood contours

By fixing two of the model parameters, and varying the other two in such a way as to keep the log likelihood at a constant value, likelihood contour plots can be produced. These provide a useful tool for examining the convergence of the log likelihood function to its maximum, and ensuring that the maximum is unique.

To produce the contour plots two parameters are first fixed. One of the two remaining parameters is then also fixed at each of a series of values covering a range over which the log likelihood contour plot is to be produced. A value of the fourth parameter that gives the desired log likelihood can then be found by the following method.

Some starting value is first chosen for the fourth parameter. The first partial derivative of the log likelihood function with respect to this parameter is calculated. The position where a tangent from this point intersects the desired log likelihood will give a new value for the fourth parameter closer to the desired log likelihood (see

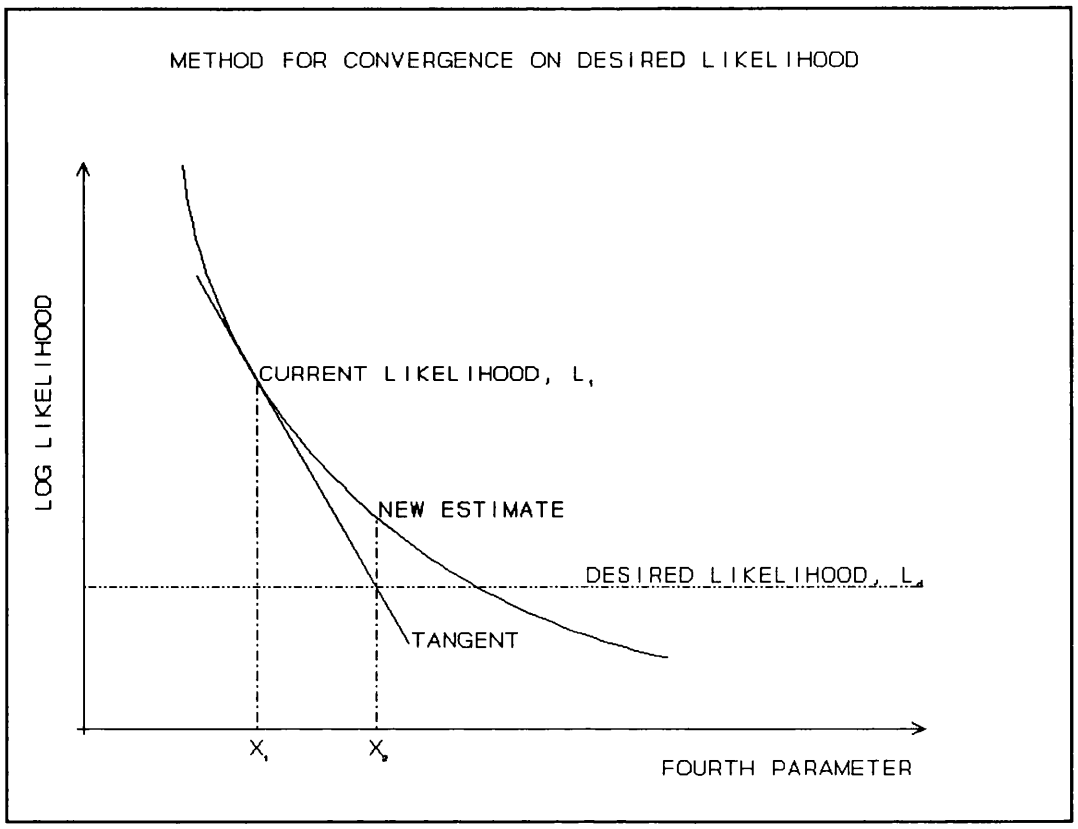


Figure 4.7 Methodology for obtaining a value of the fourth model parameter to produce a given likelihood, when the other 3 parameters are fixed.

figure 4.7). Specifically, if the current estimate is X_1, the log likelihood at this value is L_1, the first partial derivative is ∂_1 and the desired log likelihood is L_d, then a value X_2 which is closer to L_d will be found at

$$X_2 = X_1 + (L_d - L_1)/\partial_1$$

By repeating this process rapid convergence to the desired log likelihood is obtained (this consistently occurs in at most six steps, even for very inaccurate starting guesses). If there is no value for the fourth parameter which produces the desired log likelihood, this will be apparent when the series of log likelihoods which is supposed to be converging on the desired log likelihood overshoots, and starts rising again.

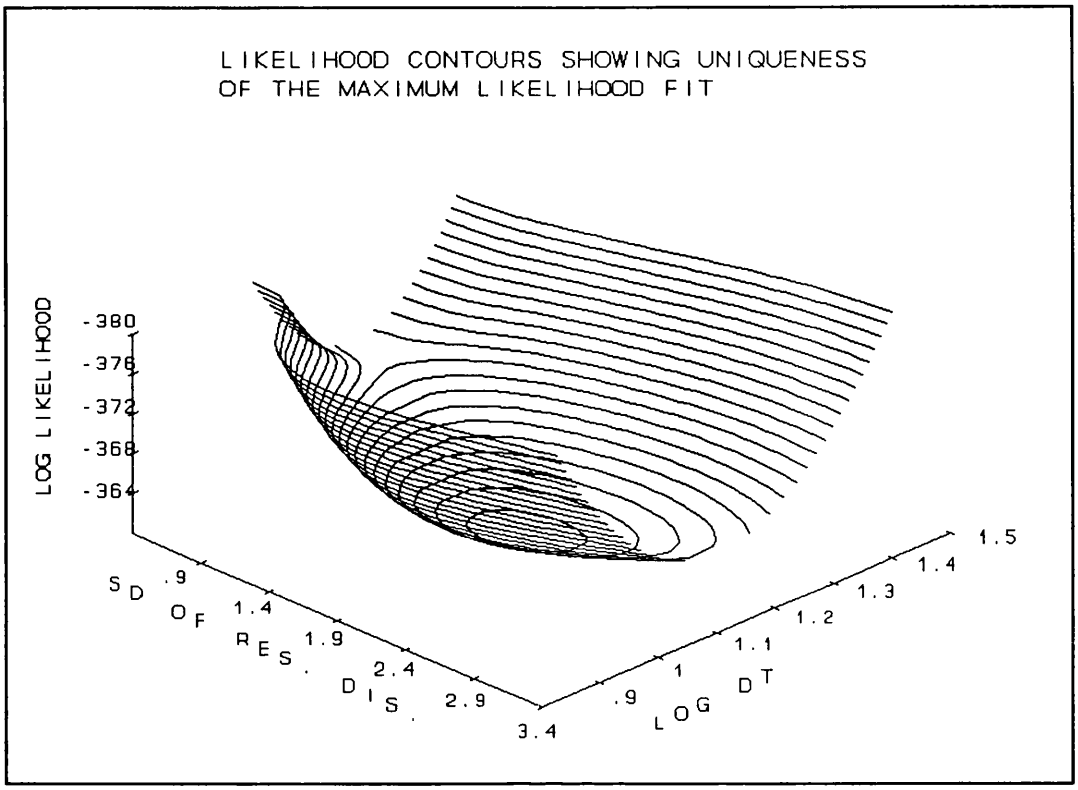


Figure 4.8 Likelihood contour plot of the SD of residual disease against the tumour doubling time for the MRC group having 1st remission between 1.5 and 2.5 years.

Likelihood contours plotted in this way show a single peak, suggesting that the maximum is unique. This seems to hold for any of the four parameters plotted against any other. The general shape of these plots also supports the independence of the parameters, since there are no long narrow peaks where pairs of values for the two parameters would give equivalent likelihoods. Some examples are given in figures 4.8 and 4.9. These are taken from the MRC AML data used as the example application in section 4.5. The negative of the log likelihood has been plotted, since this is easier to represent in three dimensions. Thus the plots show valleys or troughs rather than peaks.

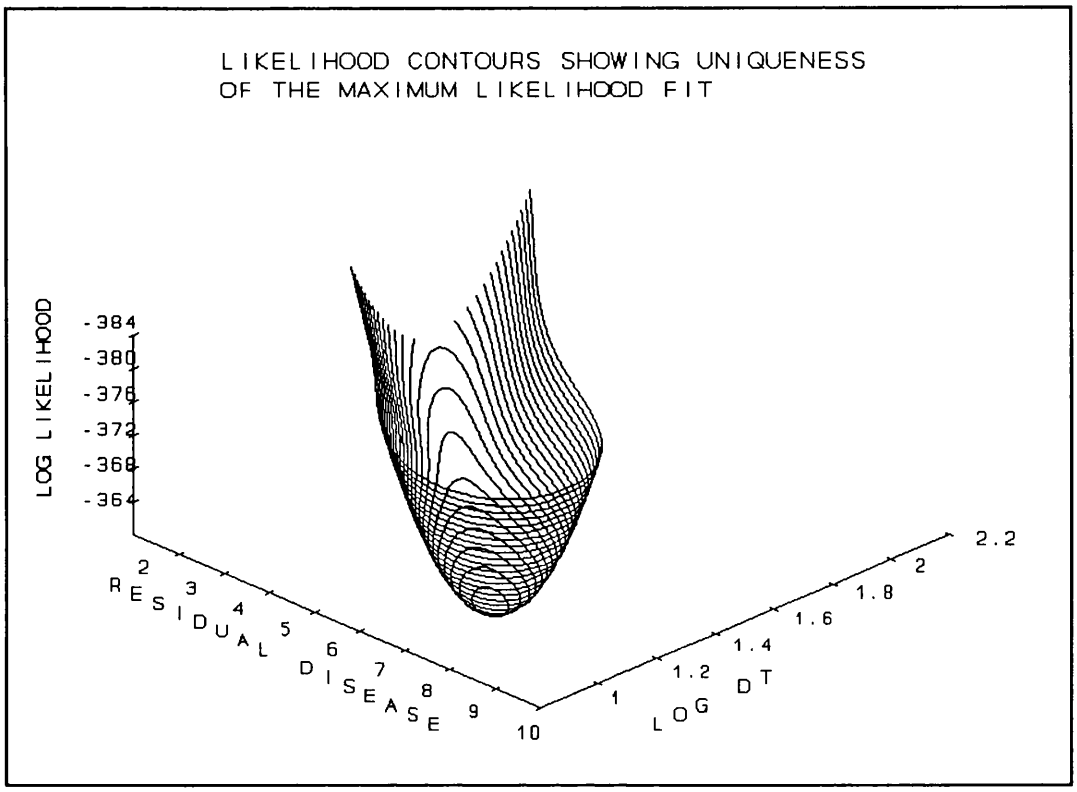


Figure 4.9 Likelihood contour plot of the residual disease against the tumour doubling time for the MRC group having 1st remission between 1.5 and 2.5 years.

4.5.2 Testing for differences in the model's estimates

When comparing different patient groups it is possible to test the significance of differences between each of the parameters of the model (i.e. the mean and standard deviation (SD) of the log of the volume of resistant tumour, and the mean and SD of the log of the doubling times). This is done by comparing the log-likelihood value for the fit of both curves with all the parameters unrestricted in value, with a similar log-likelihood value where the fit has the restriction that the parameter of interest must have the same value in both curves. This is the Likelihood Ratio Test (Silvey, 1970), and produces a chi-square statistic and p-value, which indicates the degree to which the difference between the two curves depends on that particular parameter.

4.5.3 Goodness-of-fit tests

To test whether the model provides a good fit to the observed durations of response, the latter can be divided into intervals, and a χ^2 goodness-of-fit test performed between the observed and expected numbers in each interval (Lee, 1980). Intervals are chosen to ensure at least 5 expected events in each. P values of > 0.05 suggest an adequate fit.

4.5.4 Standard errors of model estimates

Once the parameters have been estimated the second partial derivatives at the maximum likelihood parameter values can be used to estimate standard errors for each of the model's parameters. Thus, confidence intervals can be provided for the model estimates. The variance for each of the four parameters is found from the diagonal elements of the inverted Information matrix $I(\hat{\theta})$, where

$$I(\hat{\theta}) = \left(\frac{-\partial^2 \log L(\hat{\theta})}{\partial \hat{\theta}_i \, \partial \hat{\theta}_j} \right)_{4x4}$$

where $\theta = (\theta_1,\theta_2,\theta_3,\theta_4)$ are the model parameters, i.e. $\theta_1=\mu_v$, $\theta_2=\sigma_v$, $\theta_3=\mu_g$, $\theta_4=\sigma_g$.

4.6 Computational methods

4.6.1 Numerical integration methods

Equations {4.5} to {4.40} present a large number of integrals and double integrals to be evaluated. Fortunately, for all the double integrals the inner integral is identical namely $\int_{-\infty}^{U_t} N_g \, dg$. This integral has therefore only to be evaluated once for each time to enable the calculation of both the function value and all the first and second partial derivatives.

The general method of integration used was Gaussian quadrature (Davis & Polonsky, 1965). This divides the region of integration into a number of unequal intervals, where the abscissas are zeros of Legendre polynomials. The general formula is as follows:

$$\int_a^b f(x)\,dx = \frac{(b-a)}{2}\sum_{i=1}^{n} w_i f(x_i)$$

where n can be chosen to give a desired level of accuracy. The associated abscissas and weights (x_i's and w_i's) have been calculated for various values of n (Davis & Polonsky, 1965). A value for n of 32 was used in all the programs; this was found to give results accurate to 15 decimal places (this is the common accuracy of double precision variables in Fortran compilers).

Since the integrals all have the same limits, they can all be evaluated within one program loop. Furthermore, many of the integrals are similar, and require only an

additional multiplication or division, thus making the evaluation surprisingly quick (see section 4.6.2).

With the number of equations presented, it would be quite likely that an algebraic mistake would have occurred somewhere. To check that all the derivatives were correct, numerical analysis versions of the derivatives were calculated, and checked against the results from the equations given. In this way, all algebraic errors were eliminated. The technique of Richardson extrapolation (Kincaid & Cheney, 1991) was used for the numerical analysis calculations of derivatives. Two steps proved to be adequate. Thus in calculating the first derivative of P with respect to μ_v for example, P is calculated at four values of μ_v either side of, and very close to, μ_v itself; say $\mu_v-2\delta\mu_v$, $\mu_v-\delta\mu_v$ and $\mu_v+\delta\mu_v$ and $\mu_v+2\delta\mu_v$. The following formula then gives the first partial derivative:

$$\frac{\partial P}{\partial \mu_v} = \tfrac{4}{3}\,\phi(\delta\mu_v) - \tfrac{1}{3}\,\phi(2\delta\mu_v)$$

$$\text{where } \phi(\delta\mu_v) = \frac{P(\mu_v+\delta\mu_v)-P(\mu_v-\delta\mu_v)}{2\delta\mu_v}$$

It is necessary to take a value for $\delta\mu_v$ which maintains as many significant figures as possible in the difference between the four function values (at $\mu_v-2\delta\mu_v$, $\mu_v-\delta\mu_v$, $\mu_v+\delta\mu_v$, $\mu_v+2\delta\mu_v$) while being small enough to provide an accurate derivative. It proved possible to get results accurate to approximately 10 significant figures, this being the correspondence observed between the numerical analysis derivatives and the derivatives obtained from equations {4.5} to {4.40}.

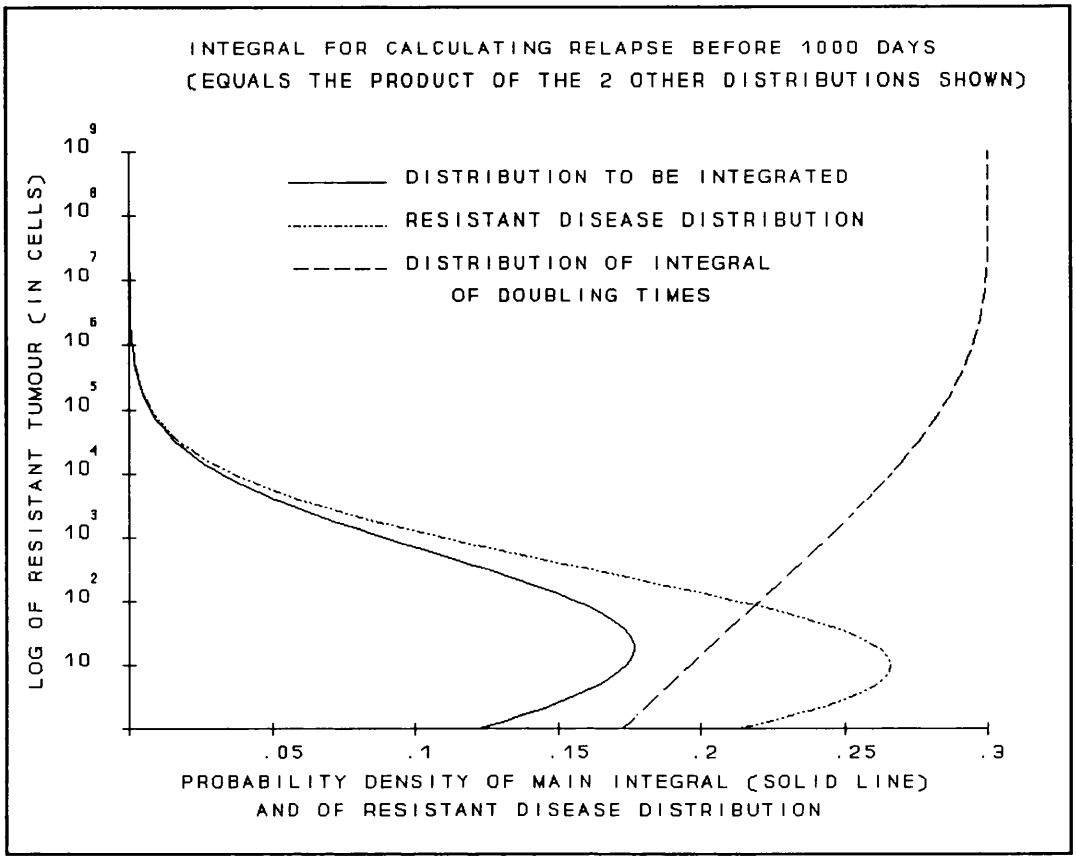


Figure 4.10 Example of the probability density function to be integrated to derive the probability of relapse before a time t. Parameters are $\mu_v=1$, $\sigma_v=1.5$, $\mu_g=30$, $\sigma_g=.25$. Probability density is scaled to 1 for the doubling times density function.

The function P is a product of a normal distribution parameter (the resistant tumour) and a normal distribution integral (from the doubling times). As such P is generally a skewed normal distribution (see figures 4.10 and 4.11) although sometimes the shape is more complex (see figure 4.12). Since the doubling time integral is always by definition less than unity, a reasonable integration range is obtained by considering a range of values either side of the mean resistant tumour volume. With the accuracy of the integration approximation being of the order of 10^{-15} a value of 8 SD's either side of the mean resistant tumour volume is satisfactory. (For a standard normal deviate, this range covers all normal distribution values greater than

10^{-15}). Some examples showing the integration method are given in figures 4.10-4.12. The two components which when multiplied produce the final function to be integrated are also shown for comparison. An example where the function becomes more complex, for small time values, is shown in figure 4.12.

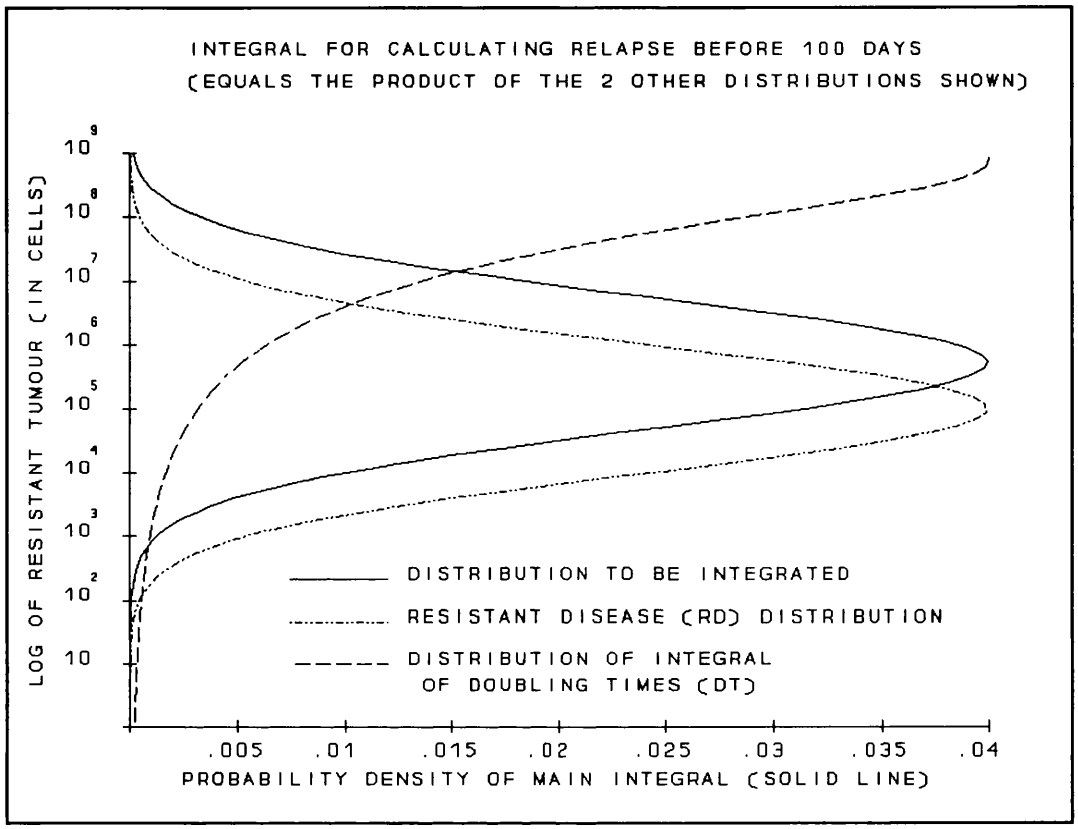


Figure 4.11 Example of the probability density function to be integrated to derive the probability of relapse before a time t (in this case 100 days). Parameters are μ_v=5, σ_v=1, μ_g=20, σ_g=.3. Probability densities are scaled to .4 & 1 for the resistant disease and doubling time density functions.

4.6.2 An example application

To demonstrate the application methodology consider one of the examples given in the next chapter, e.g. the second remission AML curve from the MRC8 trial (Rees

et al., 1986), for patients with first remissions having lasted between 1.5 and 2.5 years.

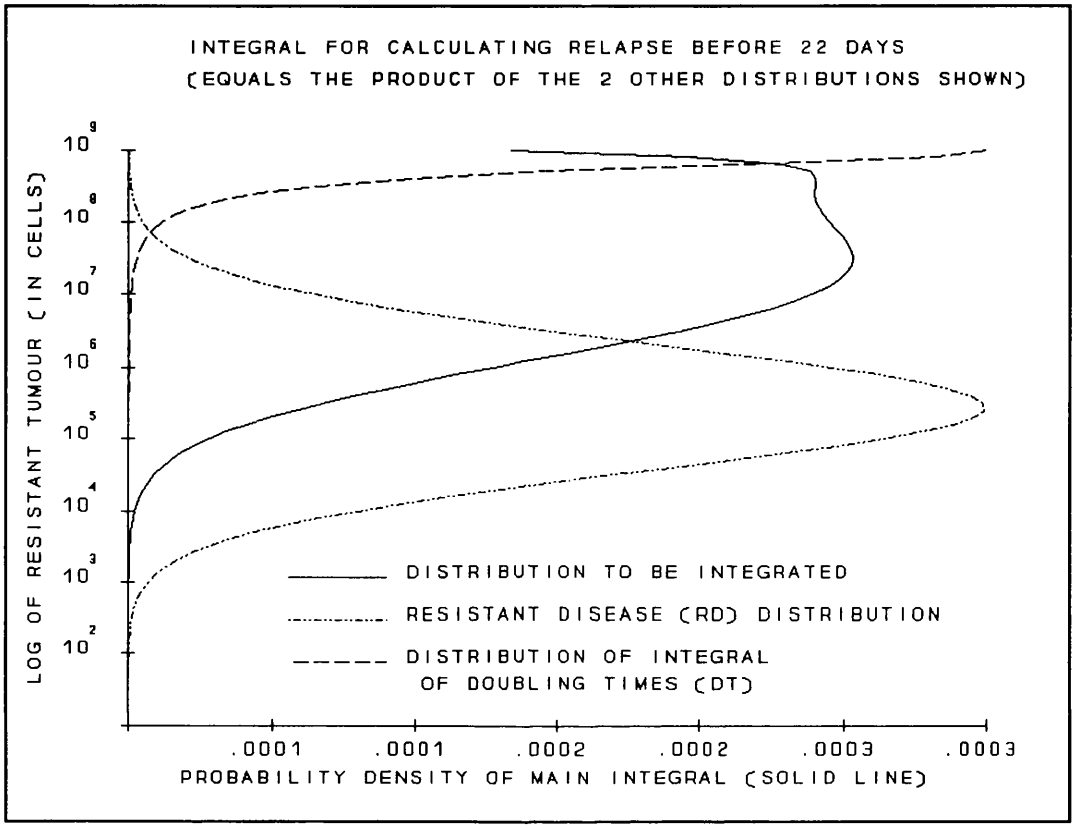


Figure 4.12 Example of the probability density function to be integrated to derive the probability of relapse before a time t (in this case 22 days). Parameters are μ_v=5, σ_v=1, μ_g=20, σ_g=.3. Probability densities are scaled to .4 & 1 for the resistant disease and doubling time density functions.

Initially, a rough guess was made for values of the four parameters which might produce a model curve similar to the actuarial curve. In this case these guesses were as follows: μ_v=2, σ_v=2, μ_g=20, σ_g=0.3. (the method is relatively insensitive to these starting guesses, unless they are very wide of the mark). The log-likelihood was evaluated for these estimates. This effectively compares the model fit (based on the four estimated parameters) to the actuarial response duration curve. The log-likelihood for this initial guess was -366.89. The first and second partial

derivatives were derived for the four parameters at these values. The first partial derivatives were -7.71, 1.74, -76.8, 27.5 for μ_v, σ_v, μ_g and σ_g respectively. The first step of Newton's method produced a new estimate based on these derivatives of $\mu_v=1.92$, $\sigma_v=1.71$, $\mu_g=13.5$, and $\sigma_g=0.249$. This procedure continued until the incremental change required in each parameter at the next step was very small ($<.000001$ of the parameter value itself). At this point the log likelihood no longer increased, and the derivatives were all close to 0 (they were in fact all $< .0000001$). The final log-likelihood was -359.40. The fit of the model to the actuarial curve is given in figure 5.3 and the final parameter values are given in Table 5.1 (see chapter 5). Running on a 20 MHz 80386 IBM compatible PC with a maths co-processor, the whole fitting procedure took just 14 seconds. Nine steps of the Newton algorithm were required.

The log-likelihoods for some of the 63 times, both censored and complete, given the initial guesses for the parameters, and then the final 'best' parameters, are shown in table 4.2.

4.7 Summary

Assuming that sufficient treatment is given to eradicate sensitive tumour cells, any residual tumour remaining will be resistant. Taking into account the assumptions previously described in this chapter, the volume of that resistant tumour will be log-normally distributed within a population. Where very short durations of treatment have been used as to make it doubtful that all anti-tumour effect has been

achieved, the model results may have to be interpreted with caution. Interpretation may also prove difficult where complicated regimens are given over long periods of time, since if long delays occur between different sections of the regimen, resistance may have been acquired in the intervening periods.

The estimates of resistant tumour volume for two different treatments given to similar patient groups (e.g. those treated in the two arms of a RCT) can be compared. The difference in tumour volume can be considered as the extra 'cell-kill' achieved with the more effective treatment.

Table 4.1

Tests of non-normality (the Shapiro-Francia statistic) for distributions resulting from simulations of the acquisition of resistance by random spontaneous mutation. Simulations were started at 1 sensitive cell, and stopped when the total population had reached the indicated number of cells. 500 separate simulations were performed at each set of parameter values.

mutation rate	proportion showing resistance	birth & death rate parameters[a] b	d	Number of cells at which simulation was halted	Shapiro-Francia statistic	P-value
10^{-4}	1.00	0.505	0.495	10^5	0.946	$<.001$
10^{-5}	0.98	0.505	0.495	10^5	0.995	.1
10^{-6}	0.29	0.505	0.495	10^5	N/A[b]	
10^{-4}	1.00	0.510	0.490	10^5	0.935	$<.001$
10^{-5}	0.95	0.510	0.490	10^5	0.992	.01
10^{-6}	0.26	0.510	0.490	10^5	N/A[b]	
10^{-6}	1.00	0.520	0.480	$5x10^6$	0.983	$<.01$

[a] the probability of replicating (birth) or dying for each cell generation.

[b] with such a small proportion showing resistance, the distribution would no longer be expected to be log-normal, and the test statistic is therefore inappropriate.

Table 4.2

A comparison of initial versus final log-likelihoods for some sample times from the worked example.

Time (days) (*=censored)	Initial log-likelihood (parameters guessed)	Final log-likelihood (parameters estimated by model)
50	-8.43	-7.45
59	-8.10	-7.13
70*	-0.01	-0.03
103	-7.16	-6.34
120	-6.97	-6.22
124*	-0.05	-0.12
155*	-0.09	-0.20
164	-6.69	-6.09
168	-6.68	-6.09
172	-6.66	-6.09
198	-6.60	-6.11
217	-6.58	-6.14
233	-6.57	-6.18
236	-6.57	-6.18
242	-6.57	-6.20
267	-6.58	-6.27
284	-6.59	-6.33
295	-6.60	-6.37
325	-6.64	-6.48
363	-6.71	-6.63
386	-6.76	-6.72
450	-6.91	-7.00
461	-6.94	-7.04
504	-7.05	-7.23
557	-7.19	-7.45
629	-7.39	-7.76
700	-7.58	-8.05
810	-7.89	-8.49
1001*	-1.36	-2.01
1250*	-1.54	-2.15
1460*	-1.63	-2.20
2198*	-1.78	-2.27

Chapter 5

APPLICATIONS OF THE REMISSION DURATION MODEL Page

Chapter 5

APPLICATIONS OF THE REMISSION DURATION MODEL

5.1 Acute myelogenous leukaemia (AML)

The first application is taken from a medical research council (MRC) trial in acute myelogenous leukaemia (MRC8) (Rees *et al.*, 1986). This large trial afforded an exceptional opportunity to verify some of the predictions and assumptions of the model. A total of 1127 patients were treated in this trial, of whom 757 achieved a complete remission (CR). Of these, 559 subsequently relapsed, and 155 then achieved a second CR. The numbers given in this paper show some differences from those given in the paper by Rees *et al* (1986), since data updated to February 1987, kindly supplied by the authors, has been used in this analysis. In addition, the patients achieving second CR after relapse from bone marrow transplants have been excluded, since it was intended to relate durations of first and second response by their quantities of resistant disease, and these patients are likely to have had different quantities of resistant disease from the others.

The large number of patients in this trial allows a detailed examination of the durations of second CR as they relate to the duration of first CR. Application of the model (as described in chapter 4) to the duration of first CR for the whole trial produced the fit shown in figure 5.1. The corresponding model estimate for the

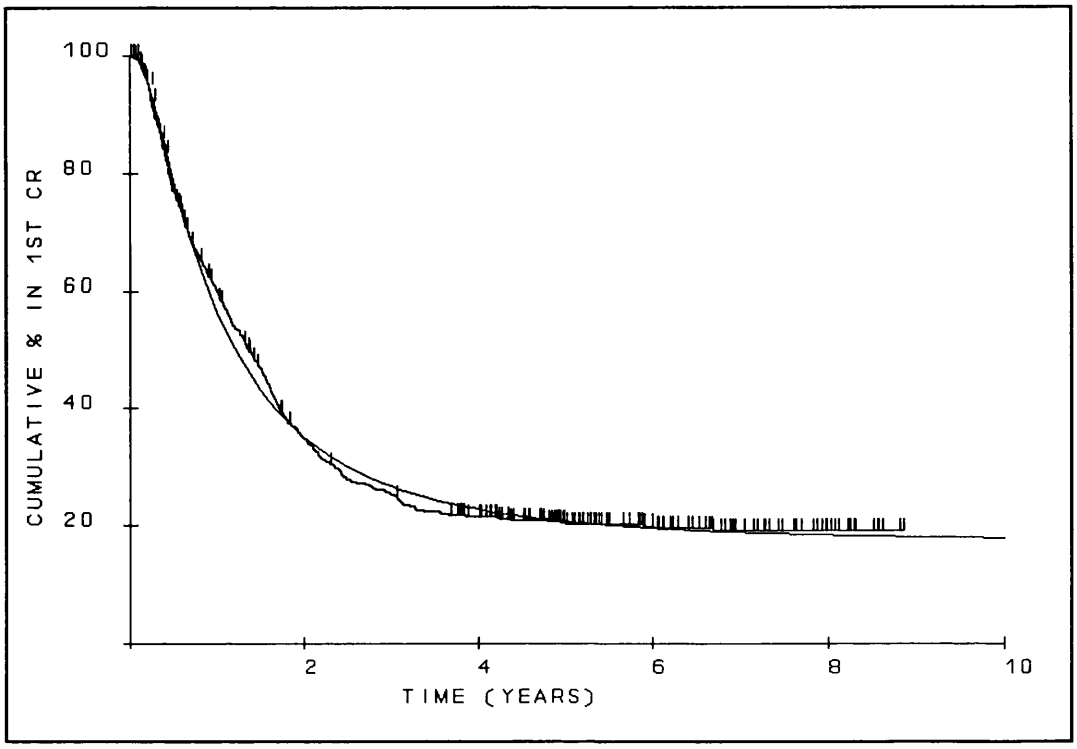


Figure 5.1 MRC AML trial 8: duration of first remission with model fit.

shape of the distribution of volumes of resistant disease is shown in figure 5.2, and the parameter values are shown in table 5.1. The lower broken line in figure 5.2 represents the "cure" threshold, as described in chapter 4. The *shape* can be plotted in this way without making any assumptions about the actual position of the "cure" threshold (see section 4.4.5). The whole picture can be completed by either assuming a particular doubling time, or by fixing the "cure" threshold at some particular number of cells, and the Y-axis can then be labelled appropriately (see, for example, figure 5.2). The parameter estimates in table 5.1 are given under the assumption that the "cure" threshold is one cell. The mean resistant tumour volume under this assumption is thus $10^{1.9}$ = 80 cells, and the estimated mean of the log of the doubling time corresponds to a doubling time of 17 days (Table 5.1). (Table 5.1 also provides standard errors (SE's) for the estimates, as described in chapter 4).

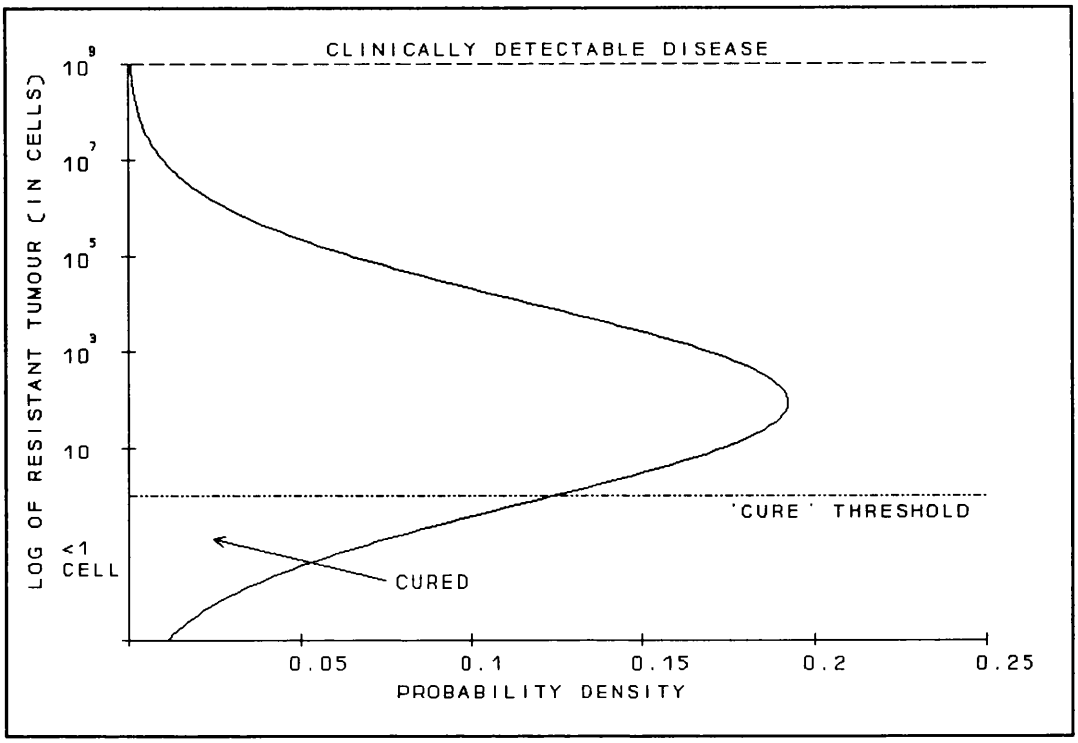


Figure 5.2 MRC AML trial 8: estimated distribution of resistant disease for first remitters.

For an individual with the mean log of the doubling time as estimated, and the lowest possible tumour volume not commensurate with cure, the response duration would be approximately 1.5 years. Thus, under the model assumptions, any patient relapsing beyond 1.5 years, would have a slower growing tumour than average. Furthermore, since the model predicts that these patients had nearly all their tumour eliminated with the first treatment, they might well have disease which is very sensitive to therapy, and thus it might be expected that they would have a better response and response duration on the second treatment (always assuming that, for an individual patient, the sensitivity and growth rate of the tumour after relapse is related to the initial sensitivity and growth rate). These effects should be accentuated for patients with even longer relapse times, for example after 2.5 years.

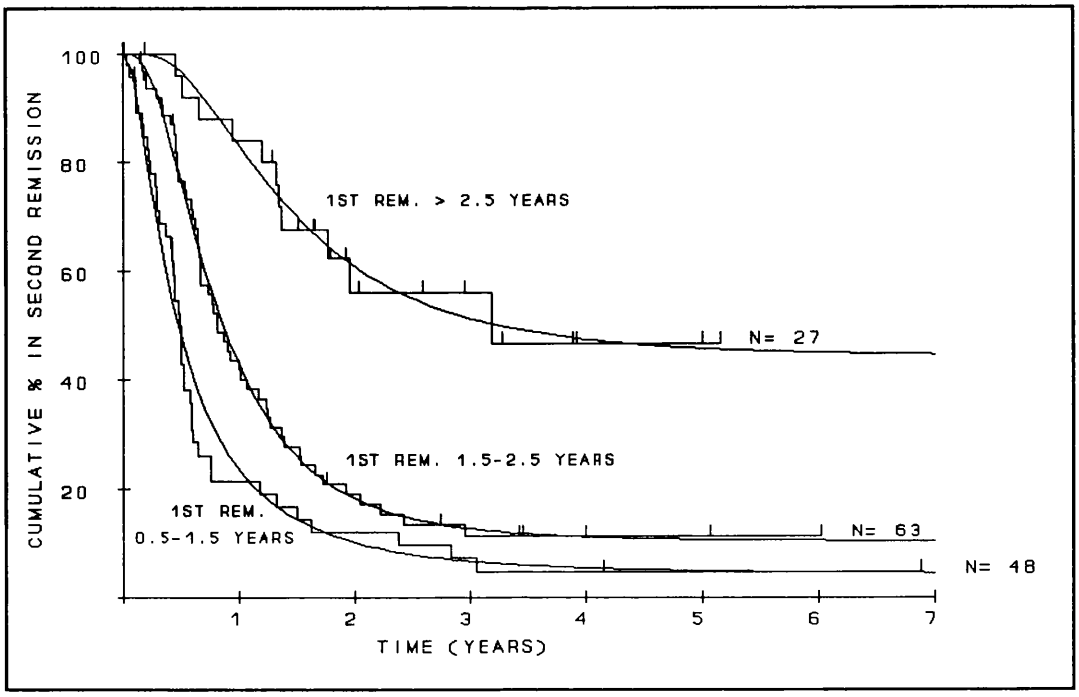


Figure 5.3 MRC AML trial 8: durations of second remission as they relate to the duration of first remission, with model fits.

The durations of second CR, as they relate to the duration of first CR (divided at 1.5 years and 2.5 years to accord with the predictions mentioned above) are given in figure 5.3 with the model fits. Very few patients with a first CR of less than 6 months achieved a second CR, and this group has therefore been excluded. The model estimates for resistant tumour and doubling times, calculated from the second remission duration curves, are given in table 5.1 and the resistant tumour estimates are shown graphically in figure 5.4. The Y-axes are fixed using the assumption that the "cure" threshold is one cell. The estimated doubling time for the patients whose first relapse occurred after more than 1.5 years is greater than the doubling time for those relapsing earlier. For patients relapsing initially after 2.5 years this trend is even more pronounced, and these patients clearly have a better second remission duration. Furthermore, the estimated resistant disease after second treatment for this

latter group is very low (mean = $10^{0.1}$ i.e. 1 cell) and narrowly spread close to the "cure" threshold. Since the choice of this group from their first remission duration was on the basis that they should, given the model estimates, have a very low volume of resistant disease, this result shows a degree of internal consistency in the model estimates. Use of the model has thus provided and confirmed tenable hypotheses regarding relationships between lengths of first and second remissions. The curves in figures 5.1 and 5.3 also demonstrate clearly the hypothesised patterns in the shapes of response duration curves discussed in chapter 4. They flatten out to plateaus, which occur at the same time post treatment (approximately 3-4 years), and the curves with steeper slopes appear to represent faster growing tumours.

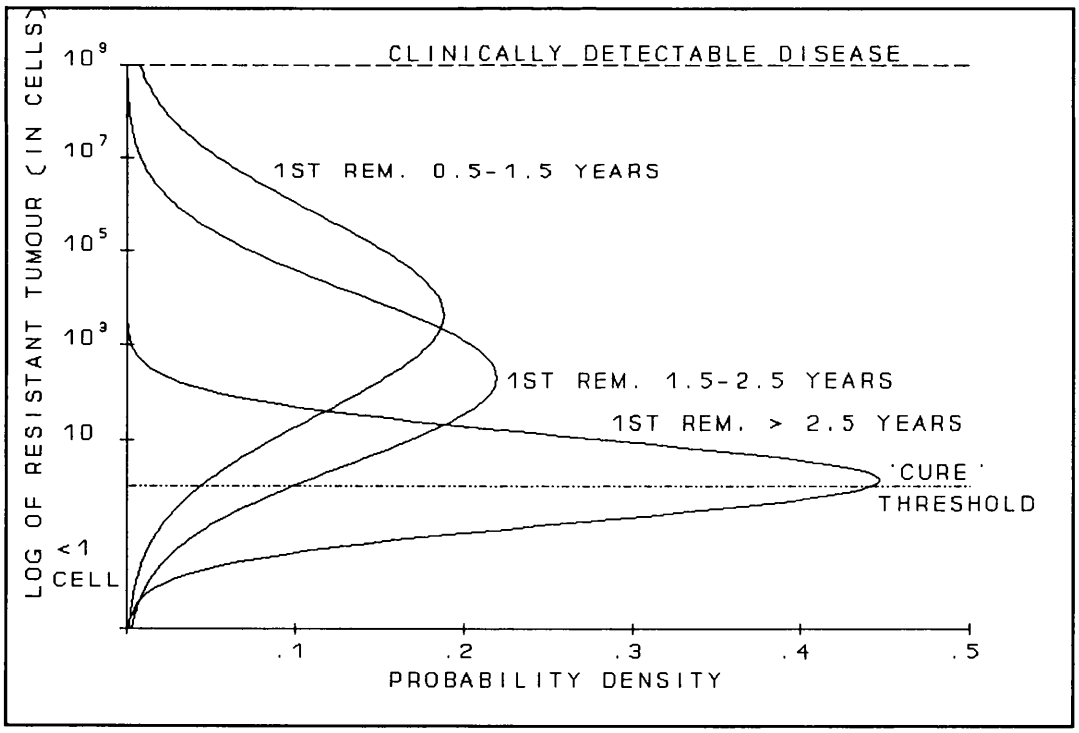


Figure 5.4 MRC AML trial 8: estimated distribution of resistant disease for the three second remission subgroups obtained from the model fits.

The four model parameter estimates for the different second CR groups were compared as described in chapter 4. The differences between individual parameters for patients whose first relapse occurred before or after 2.5 years did not achieve statistical significance, although there was a trend for the mean resistant disease to be lower in the latter group (P=0.15). The reason for the lack of statistical significance in the resistant disease estimate is interesting. Although the best fit curve for the group relapsing beyond 2.5 years is based on 40% of patients being cured, the follow-up of these patients is not sufficient to exclude the possibility that this curve will continue falling, and that most patients will subsequently relapse. (The model is still capable of producing a reasonably good fit under this assumption). If, however, the long remitters in this curve were to continue without relapsing for another year, this would no longer be the case, and the mean resistant disease estimate would be statistically significantly different from the patients whose first relapse was less than 2.5 years. This indicates the need for 'mature' actuarial data for application of the model to support these types of hypotheses. The difference in doubling time estimates was not sufficiently large to reach statistical significance, given the smallish numbers in the two groups.

For all the curves given the model fits look to be excellent. This was confirmed by performing goodness-of-fit tests as described in chapter 4. P-values of > 0.3 were produced on all four occasions.

As mentioned, different assumptions can be made for the value of the "cure" threshold without affecting either the shape of the estimated resistant disease or the

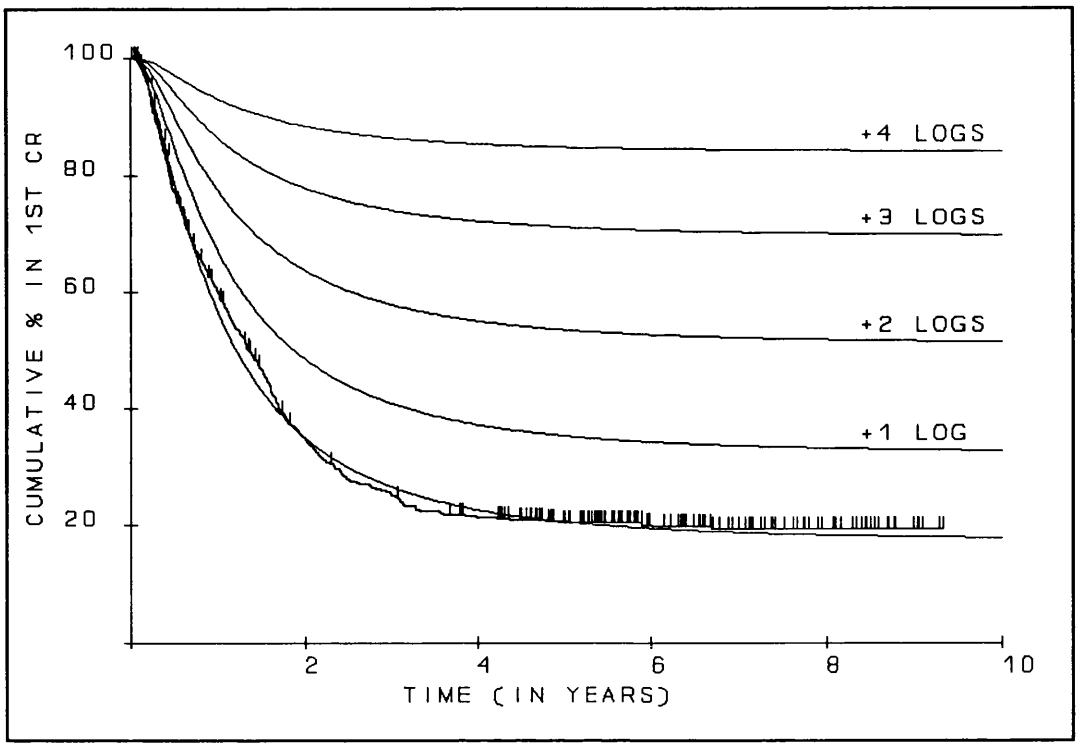


Figure 5.5 Hypothesised effect on the duration of 1st CR of systematic reductions in the log of resistant disease (in 1 log increments)

likelihood. The "cure" threshold can therefore only be gauged indirectly by examining published doubling time estimates, and seeing if they concur with those produced by the model under a particular assumed "cure" threshold value. Published estimates of doubling times in AML are rare, but in a series of 69 patients an approximate median of 10 days has been reported, (Ellison & Murphy, 1964) based on differential counts from successive bone marrow smears prior to relapse. Comparable estimates of approximately 5 days have been made in acute lymphocytic leukaemia. (Frei III & Freireich, 1965; Holland, 1968) The published estimates for AML are thus similar to the model estimates from the second remission data in MRC8, under the assumption that the "cure" threshold is one cell. This assumption for the "cure threshold" is thus supported, and lends support to the hypothesis that all leukaemia cells need to be eliminated to cure the patient.

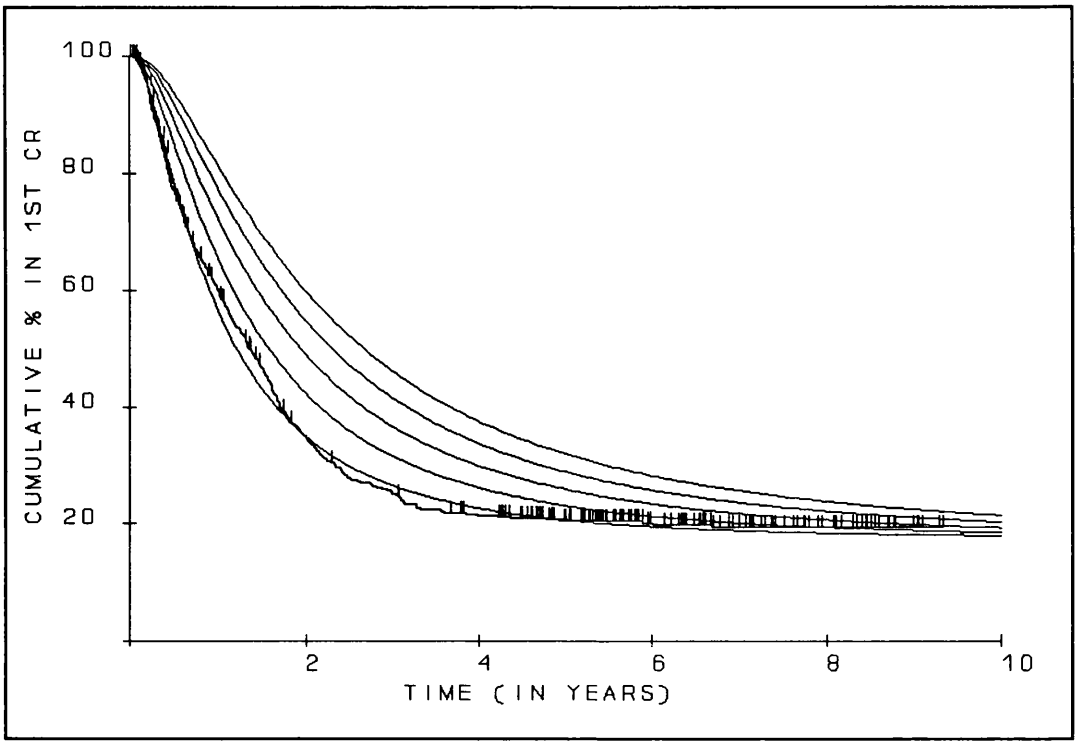


Figure 5.6 Hypothesised effect on the duration of 1st CR of systematic 5-day increases in the doubling time parameter (from 16.6 days to 36.6 days)

By making systematic changes to the parameters for the AML duration of first remission curve it is possible to demonstrate the effects that treatments with different actions should have on the curves, and thus help in the interpretation of these curves. For instance, a treatment whose sole effect was on the volume of resistant disease (effectively a treatment with different levels of cell-kill) would be expected to produce curves as shown in figure 5.5, where the main effect is on the height of the plateau, and thus on the proportion of patients whose disease is eradicated. In contrast, a treatment whose effect was to delay tumour re-growth would be expected to produce curves as shown in figure 5.6, where the effect is to cause different degrees of 'bulge' in the curve, but with no overall effect on the height of the plateau. The AML second remission duration curves show the first of these effects, and this is reflected in the parameter estimates given in table 5.1. These two

different effects should be borne in mind when interpreting curves from the applications to follow in this chapter.

5.3 Hodgkin's disease.

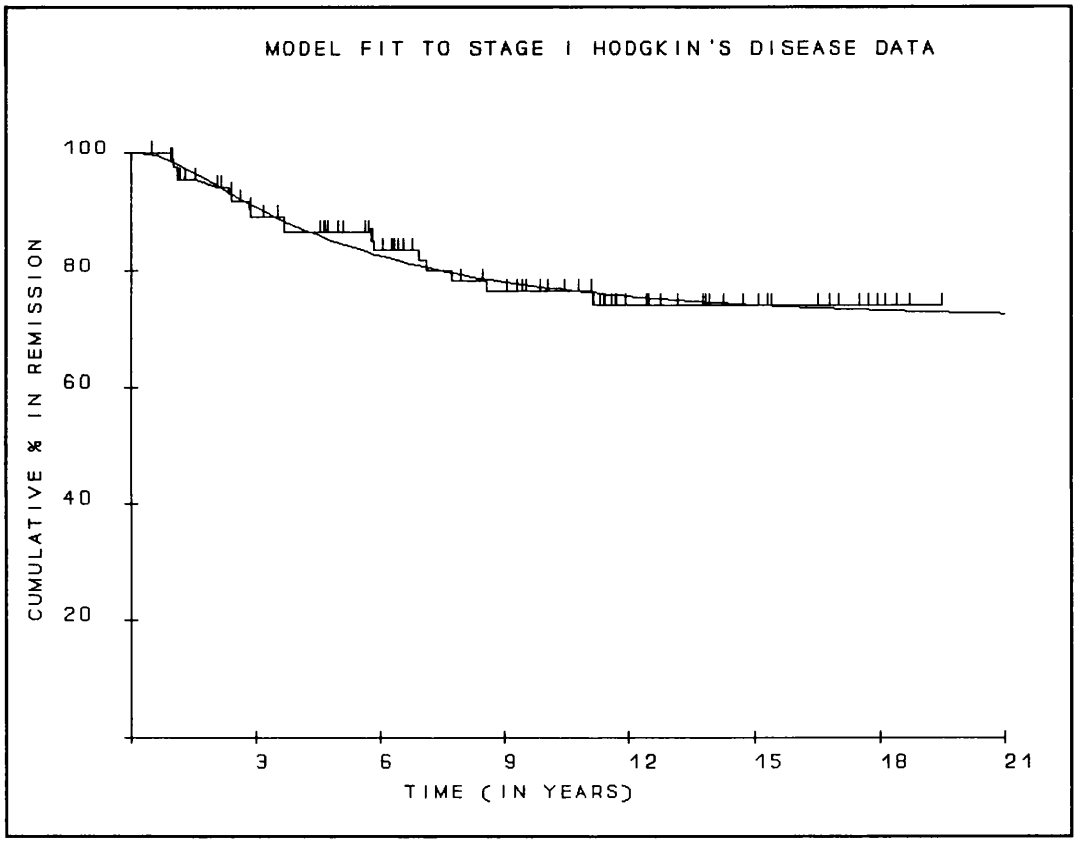


Figure 5.7 Stage I Hodgkin's disease: duration of first remission with model fit

The second application was to 89 patients with stage I Hodgkin's Disease treated at St. Bartholomew's hospital (Ganesan et al, 1990). All patients were treated with radiotherapy alone, and the 15 year disease-free rate, taken from the actuarial curve, was 74%. The actuarial curve, with the model fit, is given in figure 5.7, and the model estimate for the distribution of resistant disease is given in figure 5.8. It can be seen that, if the model estimates are correct, patients destined to relapse had a

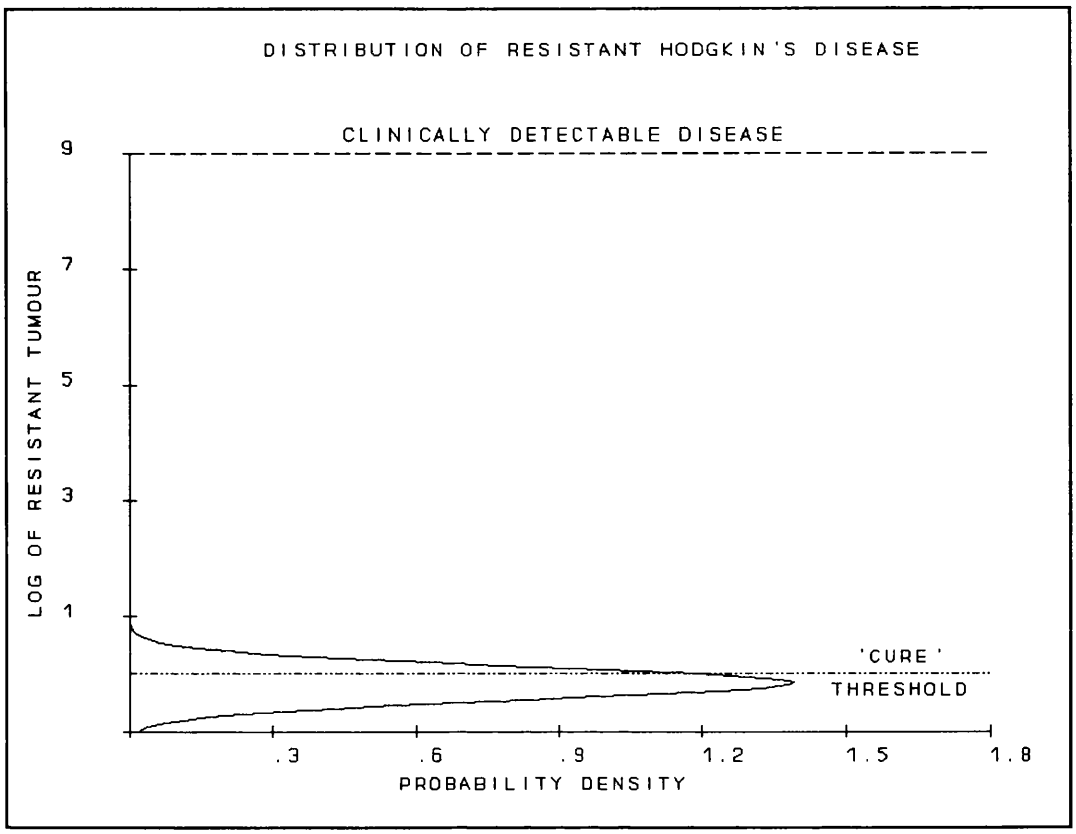


Figure 5.8 Stage I Hodgkin's disease: estimated distribution of resistant disease for first remitters

very small volume of malignant cells remaining after treatment. Six courses of combination chemotherapy (e.g. MVPP) cures the majority of later stage patients with Hodgkin's disease. If the presentation tumour volume is of the order of 10^{10} cells, a single course is likely to eradicate approximately 2 logs of disease. Thus a single course of such chemotherapy following conventional radiotherapy may well be capable of curing virtually all patients with stage I disease. Currently chemotherapy is rarely used in stage I Hodgkin's disease because of the resulting toxicity, particularly that of sterility (Waxman et al, 1987). However, one course of chemotherapy would be likely to cause very little toxicity (Waxman et al, 1987), whereas the benefits of a single course, assuming the above analysis is correct, would seem to be considerable.

To again investigate the sensitivity of the model's estimates in this case, further model fits were produced, fixing the resistant disease estimate at particular levels, and examining the resulting model fits. For example, fixing the resistant disease estimate at 3 logs the model still produced an approximate fit, as shown in figure 5.9. The fit was worse (χ^2_1=3.9, p=.05, Likelihood Ratio Test - see chapter 4),

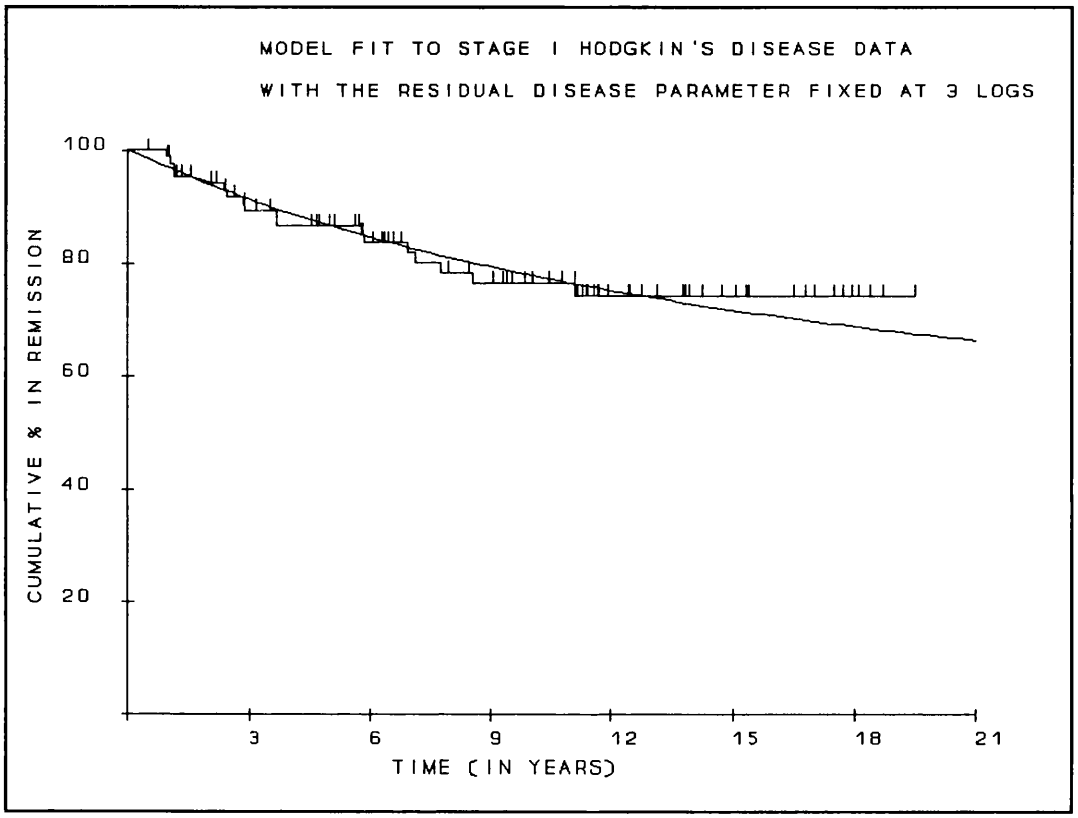


Figure 5.9 Stage I Hodgkin's disease: poor model fit with resistant disease fixed which demonstrates sensitivity to resistant disease estimate

though this only just reached statistical significance, essentially because of the few events in the curve. However, in order to obtain this fit, the plateau on the curve was lost. Although it is not absolutely certain from mere inspection that this curve will have a high plateau, other data with more patients and longer follow-up, for example the International Hodgkin's disease database (Somers *et al.*, 1990) suggests that this is extremely likely, and thus the fit is inappropriate. Under the model

assumptions it can therefore be asserted with some confidence that very little resistant disease remains after radiotherapy.

5.4 Operable breast cancer - the effect of tumour grade.

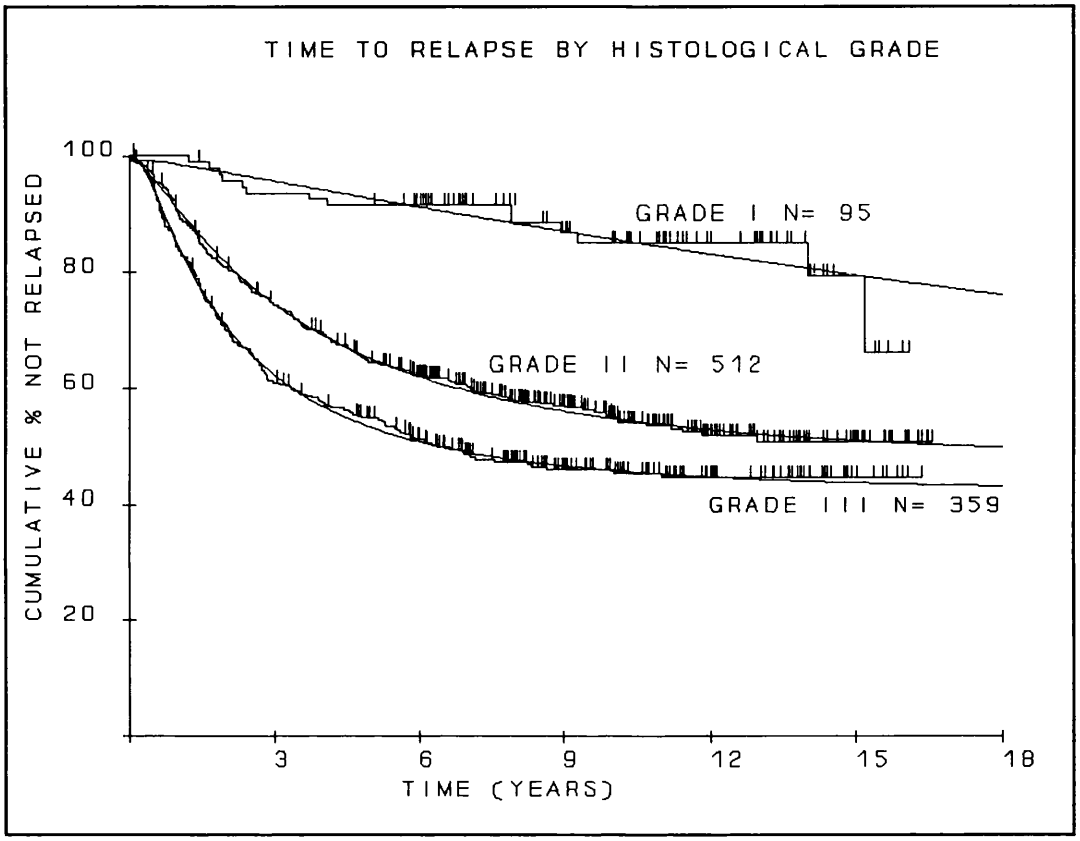


Figure 5.10 Time to relapse by tumour grade in operable breast cancer, with model fits.

As another partial validation of the model assumptions, the model was applied to a series of operable breast cancer patients with invasive ductal carcinomas from Guy's hospital for whom the patient's tumours had been graded histologically (the ductal group comprises about 80% of patients, and is the only group that can be consistently graded). Grading consists of an examination of excised tumour under the microscope to assess mitotic activity, nuclear pleomorphism and tubule formation (Bloom & Richardson, 1957). Three categories are finally produced: grade 1 (well

differentiated), grade 2 (moderately differentiated) or grade 3 (poorly differentiated). Essentially tumour grade is thought to measure the aggressiveness of the disease, with grade 3 being the most aggressive. At Guy's hospital tumour grade correlates very strongly with subsequent prognosis, both following initial curative therapy, and after relapse, and is independent of other known prognostic factors (Badwe *et al.*, 1991). Thus grading should be correlated with tumour growth rate, and the model should be able to detect and quantify this correlation.

The patients were those obtained from a computerised database at Guy's hospital between the years 1975 and 1985 (this was to ensure adequate follow-up on all patients). They were both pre and post-menopausal, with an age range of 20-70. The three actuarial curves for time to recurrence, with their associated model fits, are given in figure 5.10, and the associated parameter values are shown in table 5.2. It can be seen that the model estimates show large differences in doubling times between the three groups.

There are very few relapses in the group of patients having grade I tumours, and these model estimates should therefore be treated with some degree of caution. However, the model estimate was for a very long doubling time in these patients. Furthermore, this was not just a result of the two late relapses (see figure 5.10), since the model gave very similar estimates when these two patients were censored at their dates of relapse. These model estimates suggest that it is possible that relapses will continue to occur in this group, even if at a very slow rate.

For the patients with grade II and III tumours the model estimates showed differences in other parameters, as well as a large difference in doubling time. However, in the grade III group the standard error for the SD of resistant disease was very large (see table 5.2), and in fact an almost identically good fit could also be obtained when the SD of resistant disease was fixed to that of the grade II group (the same was not true of the doubling time estimate). A method for performing more direct comparisons of particular parameter values between groups is given in chapter 6, where this particular data set is re-examined from a slightly different perspective.

It should be noted that for operable breast cancer, treated with surgery only, the rationale described in chapter 4 for a log-normal distribution of resistant disease is lacking. However, it is not difficult to provide an alternative rationale in this case. Some authorities consider that tumour is disseminated through the lymph nodes in breast cancer (Tubiana *et al.*, 1989), hence the relevance of number of involved lymph nodes to subsequent prognosis. There is also a clear relationship between number of involved nodes and tumour size, with, in particular, small tumours (1 cm or less in size) being much more likely to be node negative than larger tumours (for instance, in the series of 966 patients just reported, 67% of tumours measuring 1cm or less were node negative, compared with 40% of tumours greater than 2 cm in clinical size). The situation is thus likely to be analogous to the random mutation rate to resistance cases already described. As the tumour increases in size, the lymph nodes become more likely to be involved, and the chances of cells becoming established in distant organs increases. Thus there may well be a random

spontaneous chance of cells becoming disseminated, which increases as the tumour grows throughout its life, resulting in a log-normal distribution of residual disease at presentation.

Local relapse can be considered in a similar fashion, with cells being more likely to reach any local tissue as the tumour increases in size, and thus be missed by surgery. In any case, there are few local recurrences without distant recurrence (6% in this series) and distant recurrence usually occurs quite rapidly after local recurrence, so local recurrence is of less importance than distant recurrence in this argument.

This application again shows the correspondence of the model results with independent methods of measuring factors related to those used in the model. This provides further confirmation that the model assumptions relate to what is happening to the patients.

5.4 Multiple myeloma

A fourth application concerns administration of very intensive chemotherapy in multiple myeloma. Complete responses were rarely seen in this disease until the recent administration of high dose melphalan. The model was applied to the complete response duration curves for 41 patients from St. Bartholomew's and the Royal Marsden hospitals treated in this way (Selby et al, 1987), at the request of the

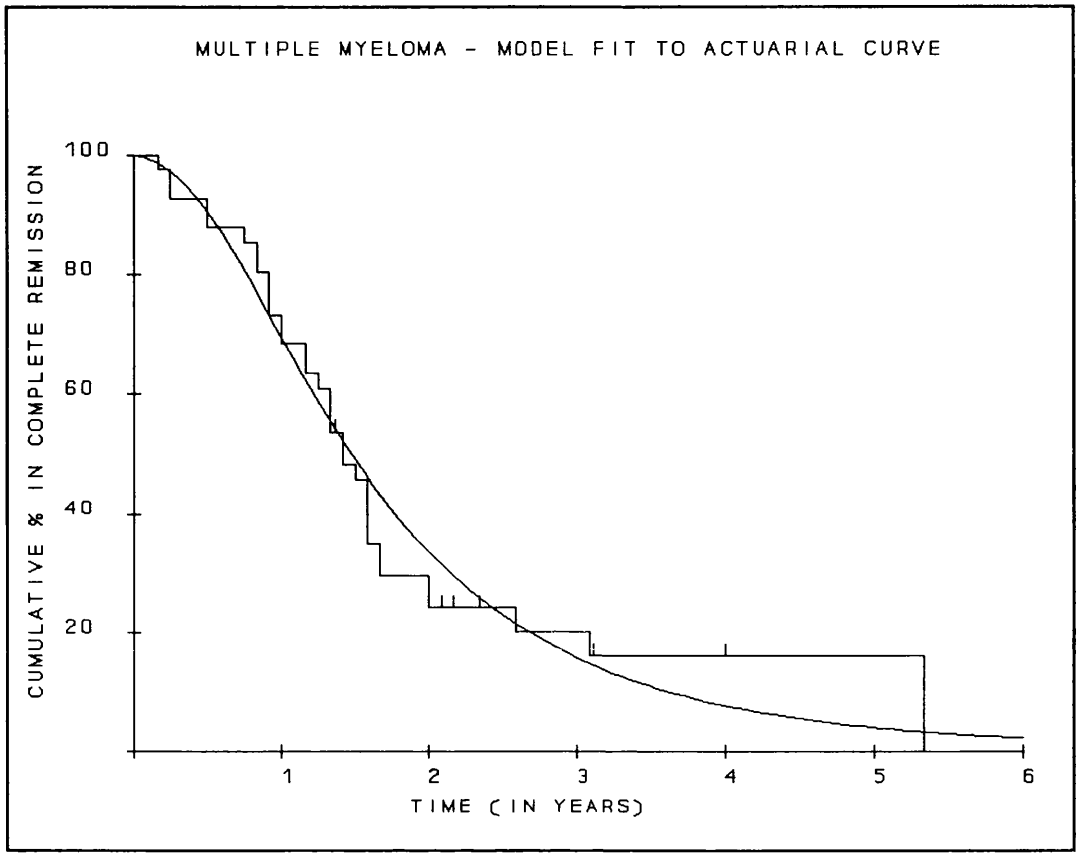


Figure 5.11 Multiple myeloma: duration of response with model fit

clinicians involved. They faced a dilemma in that although they were now seeing complete responses for the first time, it was not clear whether they should proceed by making relatively minor alterations to the current regimen, or whether more drastic changes were still required to effect cure. The model fit to the data is shown in figure 5.11, and the estimated picture of resistant disease is shown in figure 5.12. The model estimates implied that although high dose melphalan achieves a mean cell-kill of about 2-3 logs, a further 6-7 logs would be necessary, on average, to cure these patients. Thus minor alterations to the regime, such as further slight increases in melphalan dose would be unlikely to achieve cure. Additional more drastic alterations appeared to be necessary. Subsequent regimes have employed combination chemotherapy for a median of 5 courses before (late) administration of high-dose melphalan (Gore et al, 1989). Early results from these trials appear

encouraging, although longer follow-up is required to discover whether cures have been achieved. Alternative strategies involving administration of interferon after administration of the chemotherapy are also being tried, and again show considerable promise.

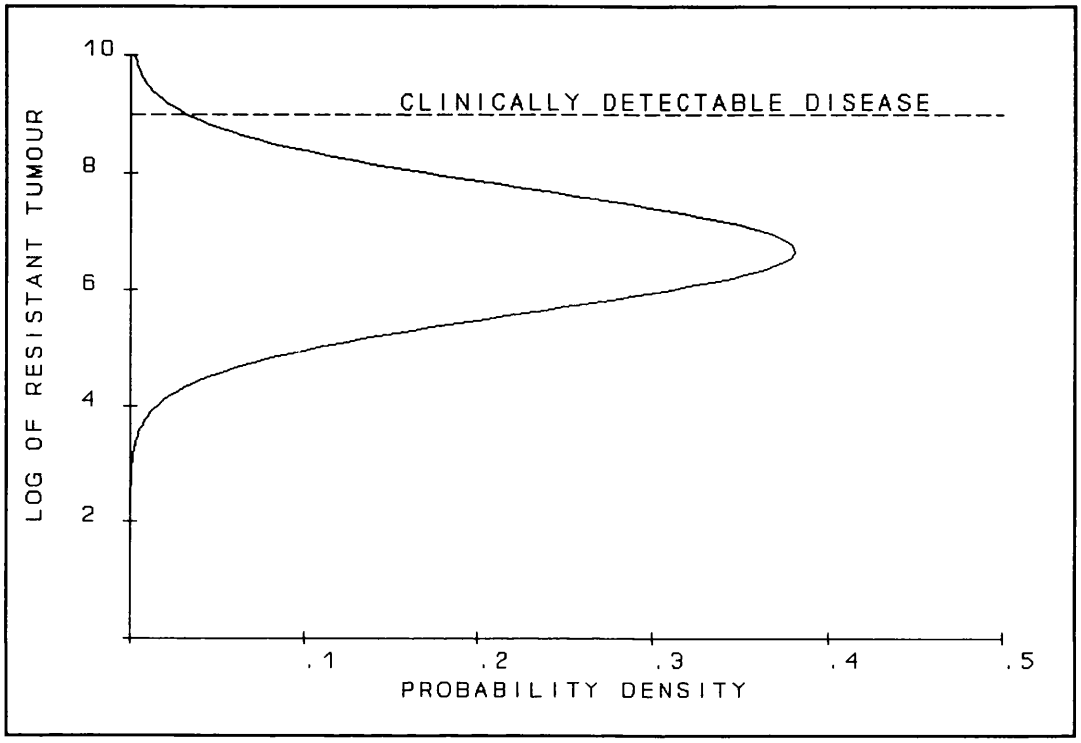


Figure 5.12 Multiple myeloma: estimated distribution of resistant disease

Since the number of patients in this group was very small, the results must be treated with caution, and the standard errors of the estimates are inevitably very wide. The sensitivity of the mean log resistant disease estimate was investigated by fixing the resistant disease at different values, and comparing the fit at these values with the unrestricted maximum likelihood fit (using the likelihood ratio test as described in chapter 4). For instance, although fixing the resistant disease at 3 logs produced a worse fit, this was not significantly worse ($p=0.36$). This again is largely a result of the small number of patients in this study.

However, the lack of a shoulder on the curve, and the fact that virtually all the patients appear to be relapsing, tends to support the inferences from the model results. The lack of a shoulder on the curve suggests that patients have not attained cell-kills much below the level of clinical detection, and the lack of long remitters (or 'cures') suggests that few if any patients had all their disease eliminated by the treatment. Thus the conclusion that more drastic measures are necessary in this situation appears to be strongly suggested by the foregoing analysis, although it is by no means proven.

5.5 Locally advanced breast cancer

The fifth application is taken from an EORTC trial for locally advanced breast cancer, and will be covered in more detail because of the variety of treatments and possible implications for therapy.

A number of different outcomes are possible for patients with breast cancer who achieve a complete remission with primary treatment. Firstly, they may remain in long term remission. Secondly, they may relapse at the site of primary disease. Thirdly, they may relapse at a site distant from their original disease. If this occurs without evidence of local recurrence it is usually assumed that micrometastatic disease was present at the time of diagnosis. Finally, both metastatic disease and local recurrence may develop.

The multicentre EORTC trial was designed to assess the contribution of cytotoxic chemotherapy and/or endocrine therapy to the primary treatment of locally advanced breast cancer by radiotherapy (Rubens *et al.*, 1989). The trial had a 2x2 design where all patients initially received radiotherapy. Patients were then randomised to receive in addition chemotherapy, hormone therapy, or both chemotherapy and hormone therapy. Although the trial can be evaluated, as explained, both in terms of local and distant recurrence, the latter is of fundamentally greater importance, since it determines subsequent survival. The model was applied to both these end-points. Application of the model to both data sets also provides an opportunity to assess possible differences in the efficacy of treatments against local and metastatic disease.

5.5.1 Patient data and treatment details

The actuarial curves analysed in this study were those from 276 patients treated according to EORTC protocol 10792. Eligibility criteria are given in Rubens *et al* (1989). A subgroup of patients who were entered into the study because of involvement of apical axillary lymph node but who did not meet the other criteria for locally advanced breast cancer has been excluded from the present analysis. All patients received radiotherapy to the breast and gland fields as their primary treatment. Patients with clinical evidence of disease progression (local and/or distant) at the end of radiotherapy were taken off study and have been excluded from this analysis. The details of the four treatments were as follows:

1. **Radiotherapy (RT)**

Radiotherapy was given to the breast and the axillary, supraclavicular, infraclavicular and ipsilateral internal mammary nodes. A dose of 4600 cGy in 23 fractions was given over a period of 5 weeks to the whole area, followed by 1400 cGy in 7 fractions to the sites of initial palpable disease.

2. **Radiotherapy and chemotherapy (RT + CT)**

After completion of radiotherapy 12 cycles of Cyclophosphamide, Methotrexate and 5 Fluorouracil (CMF) chemotherapy were given at 4 weekly intervals.

3. **Radiotherapy and hormonal therapy (RT + HT)**

Premenopausal patients received ovarian ablation (1500 cGy over 5 days) plus prednisolone 2.5 mgs three times daily for 5 years. Postmenopausal patients received tamoxifen 10 mgs twice daily for 5 years. In each case treatment was commenced within 4 weeks after the completion of radiotherapy.

4. **Radiotherapy plus chemotherapy plus hormonal therapy (RT + CT + HT)**

Radiotherapy was followed by chemotherapy and hormonal therapy as above. Hormonal therapy and chemotherapy were given concomitantly.

Time to local recurrence (or time to local progression of disease for patients with residual abnormalities in the breast following treatment) and time to development of distant metastases was recorded for each patient. In each case the response duration was measured from the time of randomisation (ie preceding the commencement of radiotherapy).

Note that none of the patients received surgery as a primary treatment, and thus the usual hypotheses concerning the distribution of resistant disease apply, in contrast to the example given earlier in operable breast cancer. For patients treated with more than one modality it will be assumed that the model estimates are of disease resistant to all the forms of treatment.

5.5.2 Model application

A. **Time to Local Progression**

The four actuarial curves for time to local progression are shown in figure 5.13. The model-derived curves fitted the actuarial curves very well for the groups treated either with RT only or with RT + CT (figure 5.14). The estimates of the mean and standard deviation of the number of resistant tumour cells are shown in Table 5.3, together with the estimated mean and range of doubling times and the results of goodness of fit tests.

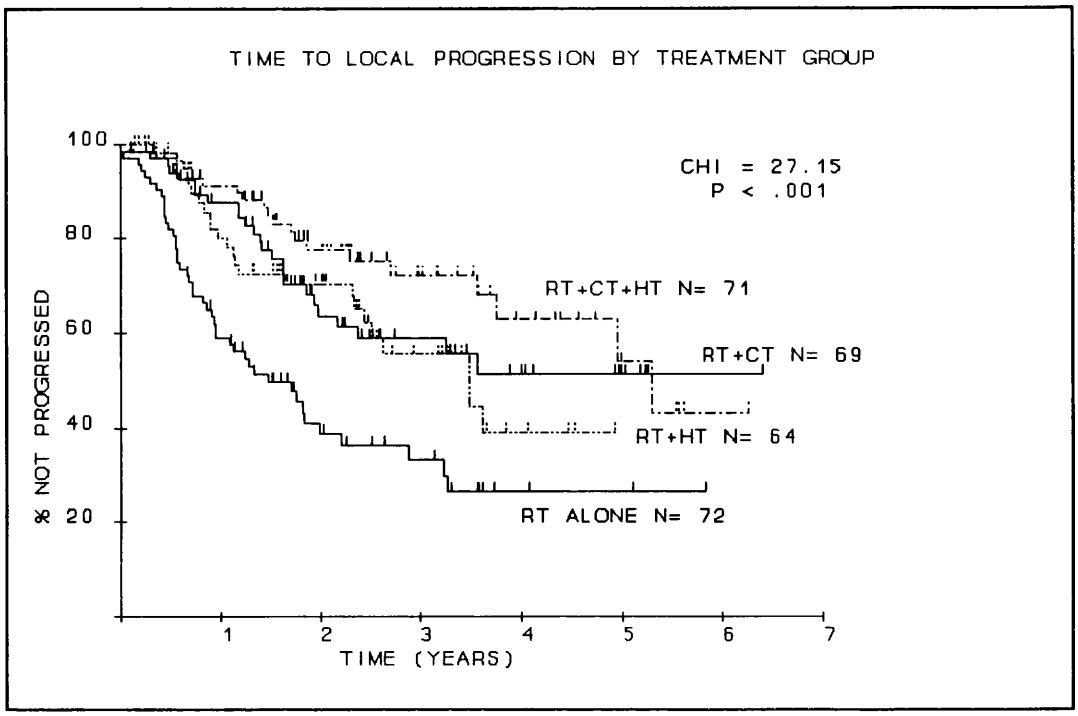


Figure 5.13 Actuarial curve of time to Local progression for the different treatment groups.

For the remaining discussion, it will be assumed that the "cure" threshold is again one cell, i.e. that all tumour cells need to be eliminated in order for the patient to be cured. This produced doubling time estimates consistent with those recorded from measurement of macroscopic disease (Shackney *et al.*, 1978).

The addition of CT showed essentially a cell-kill effect (see figure 5.5), and reduced the mean number of resistant tumour cells in the primary tumour by approximately 3 extra logs, when compared with the effect of RT alone. This corresponds with a rise in the plateau level on the actuarial response duration curves from less than 30% for the RT only group to more than 50% for RT + CT. Chemotherapy did not significantly alter the growth rate of tumours in those who relapsed (mean doubling times 25 days for RT and 24 days for RT + CT).

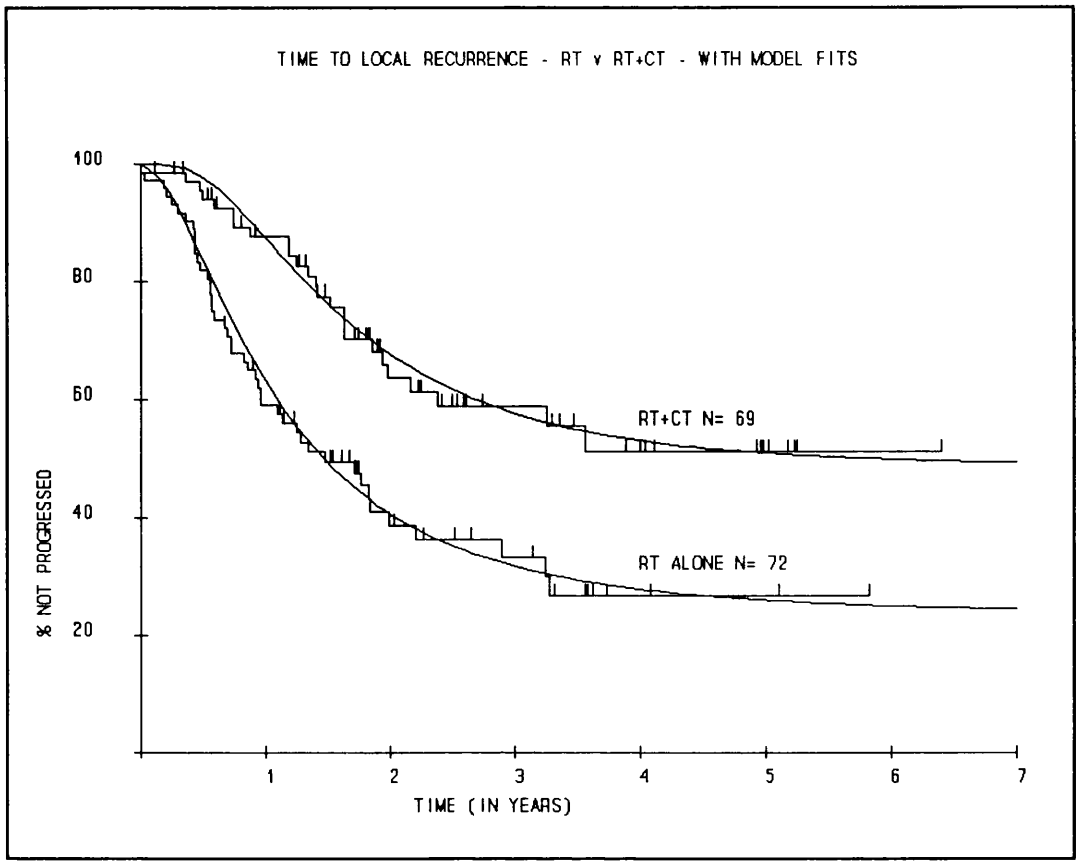


Figure 5.14 Model fits to the RT alone group and the RT+CT group for time to local progression

An alternative method of quantifying the difference in resistant disease between the two groups is given in the next chapter. Results were produced which correspond quite closely with the result obtained here.

The model failed to fit the actuarial curve for time to local progression for the 64 patients who received RT + HT (Table 5.3), Newton's method failing to converge to a maximum. An approximate fit (a range of values all gave equally good - or poor - fits) is given in figure 5.15, which demonstrates the lack of correspondence between the model fit and the actuarial curve. Further examination of the actuarial curve showed a shoulder which was not seen on the RT curve. This was followed by a slope very similar to that on the RT curve. Thirteen recurrences occurred

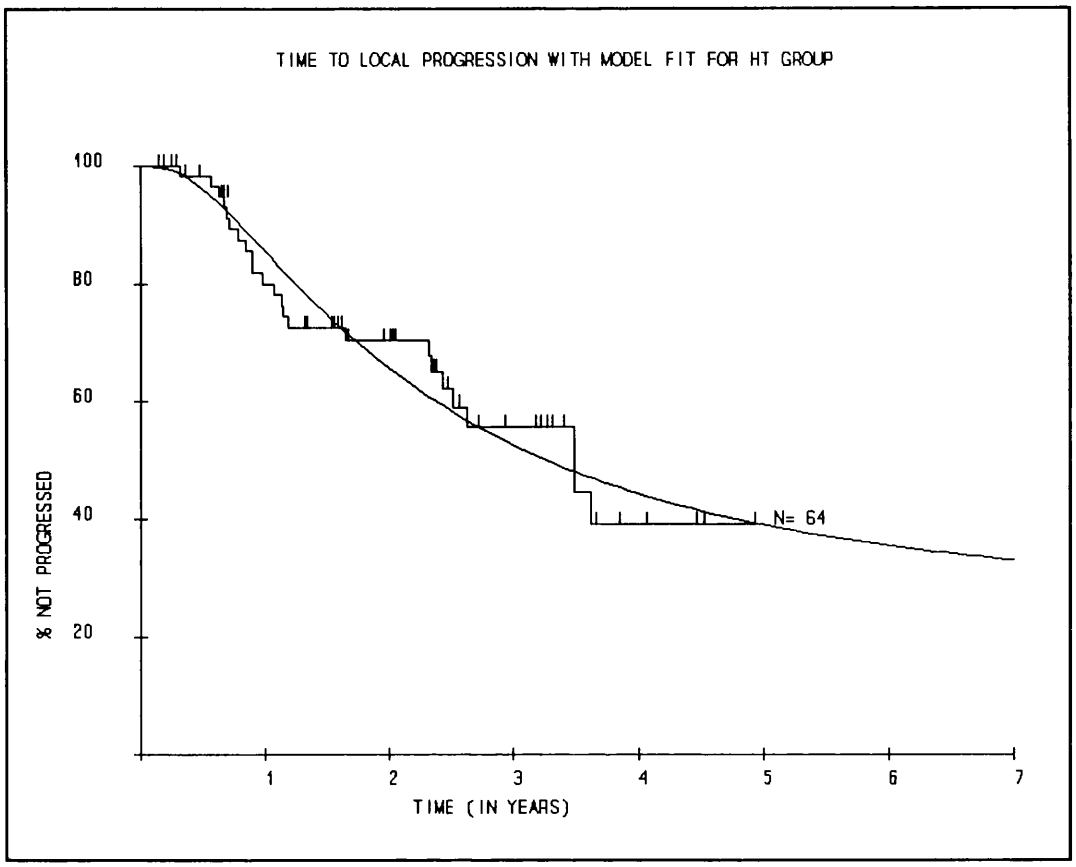


Figure 5.15 Approximate model fit to the RT+HT group for time to local progression, showing an inadequate model fit

within a 7 month period. This slope ended at approximately 15 months after which there was only a single local recurrence over a period of 14 months. This apparent plateau was followed by a second slope. By censoring times beyond 15 months, and applying the model to this censored data, a good fit was produced (figure 5.16). The prominent initial shoulder followed by a steep slope could be explained (by reference to the model) by a cell killing effect of hormonal therapy but not by an effect on growth rate. The cell kill effect of HT derived in this way was similar or slightly greater in magnitude to that of CMF chemotherapy (Table 5.3). The plateau on the curve, which relates to the cell-kill of the treatment, is of course somewhat speculative, since no events are available to fit this latter part of the curve. The pattern of recurrences in succeeding months did not, however, fit the

model-predicted curve based on this initial period of 15 months. Discussion with the clinicians involved in this study suggested that there may be a subpopulation of patients who experienced a period of growth arrest due to HT in addition to a cytotoxic effect, though this is of course impossible to prove. The secondary slope on the actuarial curve would then be explained by escape from this growth arrest and resumption of growth at the same rate as for tumours treated with RT alone. These suggestions are of course speculative in nature, particularly as the numbers of patients are small.

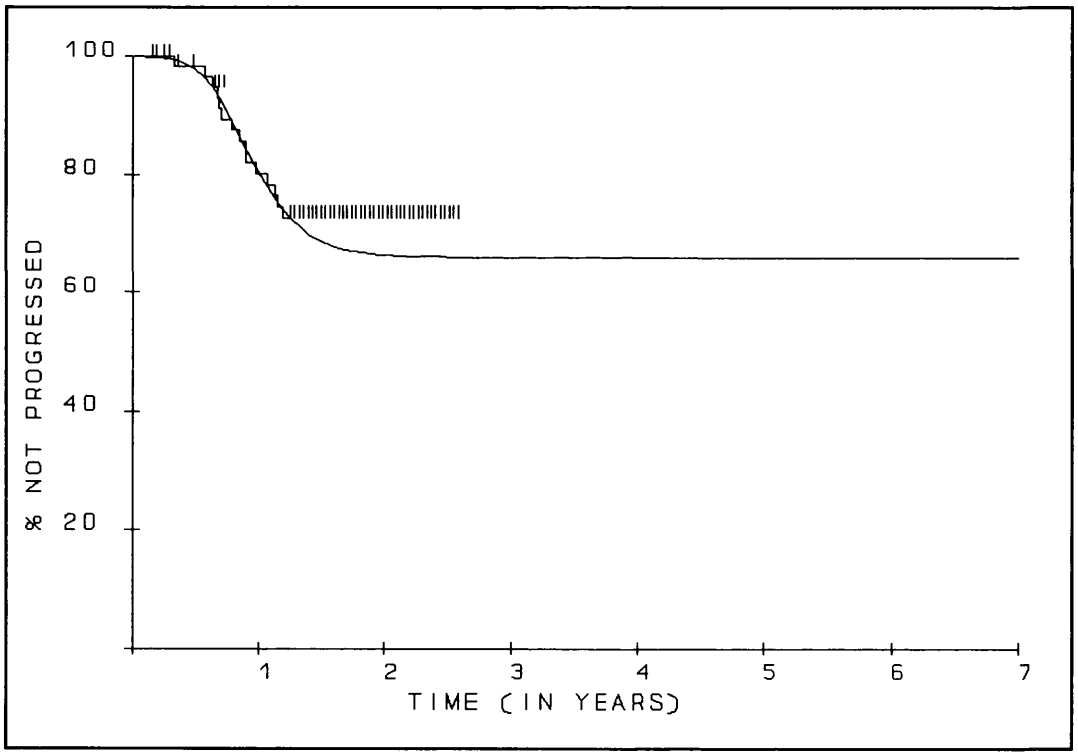


Figure 5.16 Model fit to the RT+HT group where times after 15 months have been censored, so that the model fits the early part of the curve

The model failed to fit a significant plateau to the group treated with RT+CT+HT, possibly because of two late events beyond 4 years and the relatively few numbers of relapses in this group. A number of very different model fits to this data are

therefore possible, and any model estimates produced would probably not be reliable.

B. Time to Distant Metastases

The four actuarial curves for time to distant metastases are shown in figure 5.17. There are no statistically significant differences between the curves, although the RT alone curve lies just below the other curves (p=0.1, log-rank test for RT alone v rest).

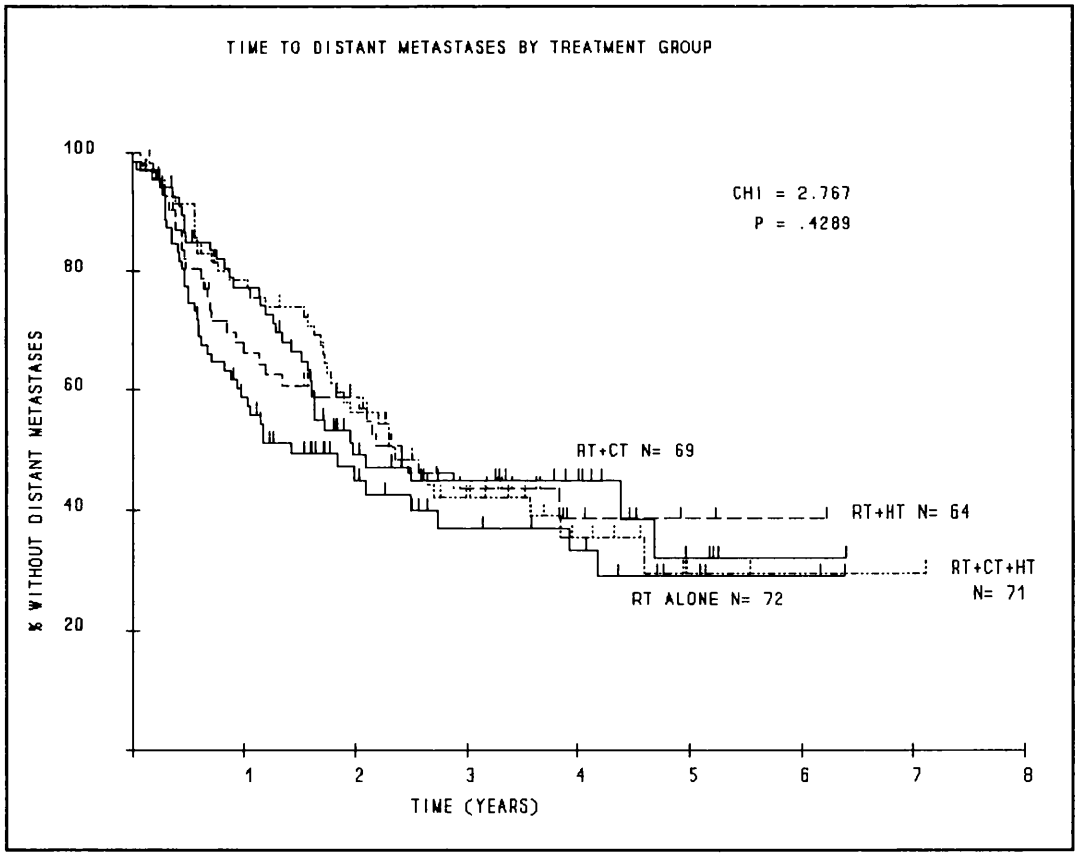


Figure 5.17 Actuarial curve of time to occurrence of distant metastases for the different treatment groups.

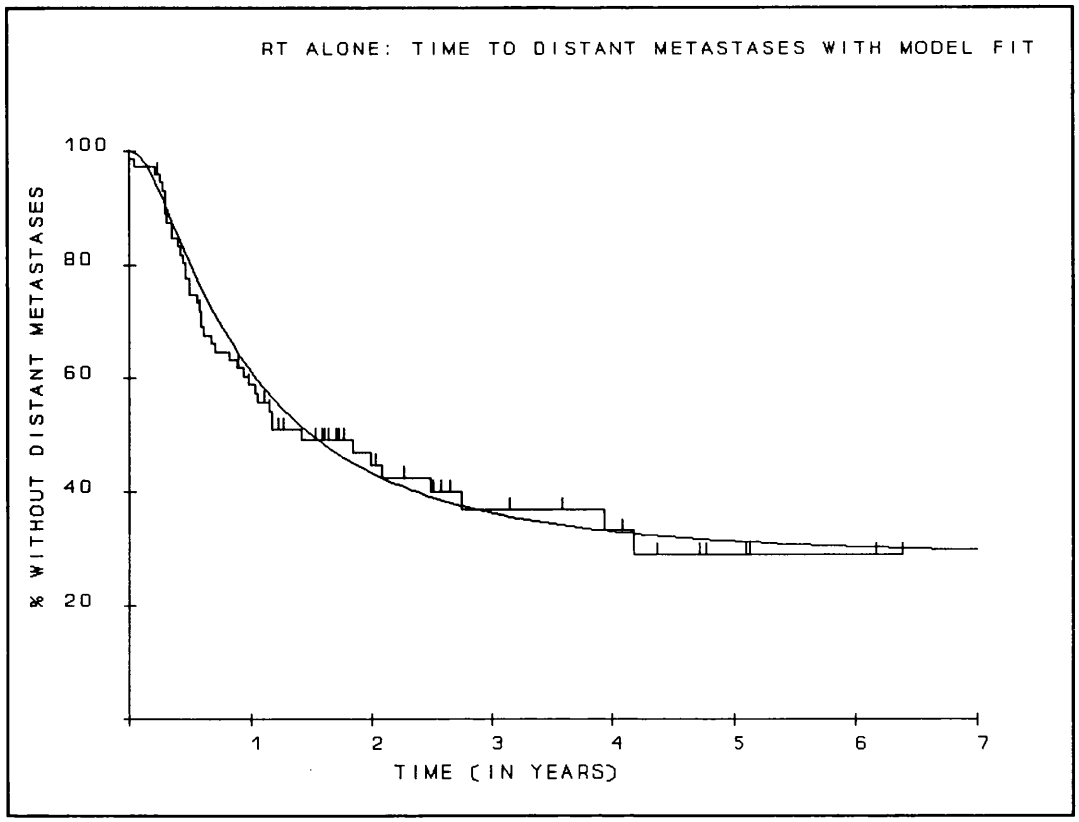


Figure 5.18 Model fit to the RT alone group for time to distant metastases

The model-derived curves fitted the actuarial curves adequately for all but the RT+HT treatment group (as an example, see figure 5.18 which gives the model fit to the RT only group). Again the Newton algorithm failed to converge to a maximum in the RT+HT group. There was a stepwise nature to the curve for this group, although this was not as pronounced as for local recurrence. The parameters for resistant tumour cell number and doubling time for the different groups are shown in Table 5.4. The resistant disease estimates are similar for the different treatment groups. Any minor differences are not statistically significant (using the Likelihood Ratio Test), i.e. similarly good fits could be obtained for each of the four curves with the assumption that they all had the same volume of resistant disease. Longer follow-up may be required however to be sure of the true height of the plateau on the curves, and thus of the resistant disease estimates. In view of the

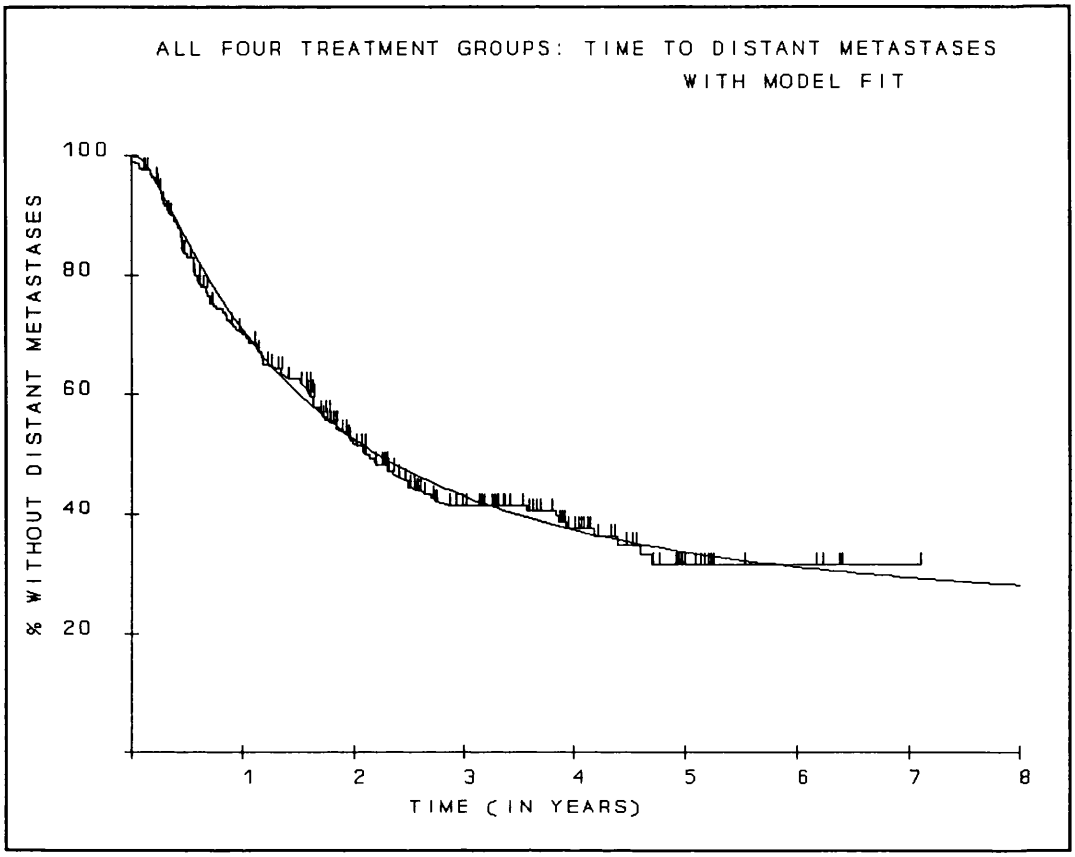


Figure 5.19 Model fit to all four treatment groups combined for time to distant metastases

similar nature of the curves for distant metastases the four groups were combined, and the model was fitted to the whole data set. This enabled more reliable estimates to be produced for the model parameters for metastatic disease. These estimates are given in table 5.4, and the model fit is given in figure 5.18. A value of under 2 logs of resistant metastatic disease was produced, with a doubling time of 24 days.

As patients who were treated with RT alone received no systemic therapy, the resistant tumour cell number for this group provides an estimate of the micrometastatic tumour volume in untreated patients. Thus, following the preceding comments, it appears that neither CT, nor HT, nor CT + HT significantly reduced

the micrometastatic tumour burden. This is in keeping with the disappointing results observed in this trial in terms of prevention of the development of metastases.

The subclinical metastatic growth rate estimate for the whole patient population (24 days) was similar to the estimates for local recurrence in the RT and RT+CT groups (24-25 days). However, RT+CT+HT resulted in a slight lengthening of the estimated mean doubling time for both local and micrometastatic disease (88 and 47 days respectively).

5.5.3 Interpretation of the results of the model application to locally advanced breast cancer

This application was encouraging in that the mathematical model also gave good fits to most of the remission duration curves for a second group of patients with breast cancer, but using different treatment modalities. The exception was in the use of hormone therapy. This suggests that the model assumptions of log-normal distribution of resistant disease, exponential growth rate of subclinical resistant disease and log-normal distribution of growth rates may have widespread applicability, although alternative assumptions may be necessary for modelling the effects of hormone therapy.

Breast cancer is usually considered to be a more slowly growing tumour than acute leukaemia. It is perhaps surprising, therefore, that the mean doubling times for

breast cancers estimated in the current study for patients receiving RT +/- CT (approximately 24 days) were only slightly longer than those previously estimated for AML (see table 5.1). However, the median duration of remission in the AML study (16 months) was similar to the median response durations for patients treated with RT in the current study. Two factors should be remembered. Firstly, the current study only concerns patients with locally advanced breast cancer and it is possible that the growth rates of tumours in such patients may not be the same as those for patients who present with operable disease (see the growth rate estimates in table 5.2). Secondly, the estimates for doubling time assume exponential growth. While this may be justified for subclinical growth, tumour growth after clinical relapse almost certainly follows Gompertzian kinetics (Norton & Simon, 1986). Most estimates of growth rate in breast cancer have been made from serial measurements of volume in clinically detectable metastases. If a single viable cell remains after treatment, approximately 30 doubling times are required before relapse from this single initiating call becomes detectable. From the results of this study, it can be calculated that, for a patient who has only a single resistant cell and has a tumour with a doubling time at the upper limit of the 2 standard deviation range (100 days), relapse would occur approximately 8 years after treatment. This is consistent with the relapse pattern observed.

Both chemotherapy and hormone therapy had a marked effect on local tumour cell kill in addition to that given by radiotherapy. Neither treatment, however, had a significant effect on the micrometastatic tumour burden (which would not, of course, be affected by the radiotherapy, being outside the radiotherapy field). Can these

findings be explained on the basis of existing knowledge of tumour kinetics? One possibility is a synergistic effect between RT and either CT or HT on loco-regional cell-kill. Tubiana et al have shown that the thymidine labelling index of tumours temporarily falls in response to RT (Tubiana *et al.*, 1989). This is followed by a period in which the labelling index increases above pretreatment levels. As CT has a greater effect on dividing than on resting cells, the period immediately after RT may be the optimal time for giving cytotoxic agents to improve local control. This recruitment effect of RT would not be expected to apply to micrometastatic cells, since they may well be outside the radiotherapy field, which may explain the lack of a significant effect of CT in these sites.

The model estimates for mean micrometastatic tumour burden were on average just less than 2 logs, or 100 cells. Upon discussion with clinicians involved in this study, this was found to be considerably lower than expected. They subsequently raised the possibility of giving moderate doses of total body radiotherapy (TBI) to these patients, with a view to eradicating this micrometastatic tumour burden, and perhaps curing more patients. Previously they had considered that the micrometastatic tumour burden would be too large to make this treatment feasible. The rationale for the approach is that TBI may be non-cross-resistant with CT or HT. The evidence, albeit tenuous, that the doses of RT given in TBI could be effective against micrometastatic disease comes from studies of spinal metastases in patients who had received internal mammary chain RT. A scatter dose of approximately 1000 Rads was given to the T3-T7 vertebrae. It was observed that patients developed less

thoracic spine metastases than had been expected (Grimard *et al.*, 1988). Thus the model generated hypotheses that resulted in suggestions for new treatments.

The poor fit of the model-derived curve to the actuarial curve for local disease progression in patients treated with RT + HT may paradoxically have led to other interesting findings. The model assumption of a log normal distribution of growth rates may well be invalid for this group. In the context of advanced breast cancer, it is widely recognised that patients with oestrogen receptor (ER) positive tumours are more likely to respond to hormonal manipulation than those who are ER negative. The number of cases in the present study for whom such measurements were available precluded application of the model to the separate subgroups. However, application of the model to separate intervals on the HT actuarial response duration curve and discussion with the clinicians involved in the study suggested that hormonal therapy had two distinct effects: first a 'cytotoxic' effect, and second a 'bimodal' effect on tumour growth. One subgroup apparently had growth characteristics similar to those treated with RT alone. The other subgroup apparently experienced a period of growth arrest followed by exponential growth, also at a similar rate to those who received RT alone.

The distribution of local resistant tumour volumes for the different groups indicate that even a small increase in tumour cell kill (perhaps with the use of different drug regimens) should have a marked effect on the long term control of local disease. Further discussion with the clinicians involved in this study suggested that the failure of CT and HT to act synergistically either in the control of local disease or in the

eradication of micrometastases could possibly have been due to the scheduling used in this study. HT could potentially have caused growth arrest and thereby reduced the effect of CT.

In conclusion, application of the mathematical model to breast cancer has led to hypotheses concerning the mechanisms of action of the different treatment modalities which would not otherwise have been considered. This has given a different viewpoint on the results of the trial, suggesting possibilities for the reorganisation of the treatment required, for instance rescheduling of the treatments, and raising hypotheses which should be considered in the design of subsequent studies.

It is noteworthy that RCTs provide very good data sets for model applications, as can be seen from this application.

A further application to this data is given in the next chapter, which examines multivariate methodologies for the remission duration model.

Table 5.1

Parameter estimates for MRC AML trial 8, assuming 10^9 cells are clinically detectable, and all these cells need to be eliminated to cure the patient.

Group	No. in group	log of remaining resistant tumour: mean (μ_v)	log of remaining resistant tumour: SD (σ_v)	doubling times: DT in days[a] (μ_g)	doubling times: SD[b] (σ_g)	95% range of DT
1st CR	751	1.9 (.39)	2.1 (.42)	17 (.05)	0.40 (.02)	3-100
2nd CR[c](1)	48	3.6 (1.5)	2.1 (.87)	10 (.17)	0.36 (.10)	2-52
2nd CR[c](2)	63	2.3 (1.4)	1.8 (1.1)	14 (.14)	0.29 (.07)	4-50
2nd CR[c](3)	27	0.1 (4.1)	0.9 (24)	19 (1.7)	0.29 (.11)	5-69

[a] converted from mean log DT, i.e. = $10^{(\text{mean log DT})}$

[b] standard deviation of $\log_{10}(DT)$

[c] the three 2nd CR groups are determined by the duration of 1st CR : group 1 = 1st CR 0.5-1.5 years, group 2 = 1st CR 1.5-2.5 years, group 3 = 1st CR > 2.5 years

(figures in parentheses are standard errors of estimates, and thus indicate the precision of the parameter estimates. The doubling time (μ_g) value is the standard error of $\log_{10}(DT)$)

Table 5.2

Parameter estimates for recurrence by tumour grade in operable breast cancer, assuming 10^9 cells are clinically detectable, and all these cells need to be eliminated to cure the patient.

Group	No. in group	log of remaining resistant tumour: mean (μ_v)	log of remaining resistant tumour: SD (σ_v)	doubling times: DT in days[a] (μ_g)	doubling times: SD[b] (σ_g)	95% range of DT	GOF[c] Test P-value
Grade I	95	7.1 (6.5)	3.6 (8.6)	1971 (1.2)	.47 (.35)	233-17000	.53
Grade II	512	0.7 (0.6)	5.0 (1.3)	71 (.08)	.43 (.06)	10-493	.33
Grade III	359	0.1 (3.6)	0.6 (18)	26 (.8)	.47 (.05)	3-215	.64

[a] converted from mean log DT, i.e. = $10^{(\text{mean log DT})}$

[b] standard deviation of $\log_{10}(DT)$

[c] goodness-of-fit

[d] times beyond 15 months were censored to produce this model fit

(figures in parentheses are standard errors of estimates, and thus indicate the precision of the parameter estimates. The doubling time (μ_g) value is the standard error of $\log_{10}(DT)$)

Table 5.3

Parameter estimates for local recurrence/progression in the EORTC locally advanced breast cancer trial, assuming 10^9 cells are clinically detectable, and all these cells need to be eliminated to eradicate the local disease.

Group	No. in group	log of remaining resistant tumour mean (μ_v)	SD (σ_v)	doubling times DT in days[a] (μ_g)	SD[b] (σ_g)	95% range of DT	GOF[c] Test P-value
RT alone	72	3.1 (1.6)	4.0 (1.5)	25 (.12)	0.29 (.09)	7-94	.31
RT + CT	69	0.1 (.44)	1.8 (4.2)	24 (.23)	0.28 (.10)	7-85	.15
RT + HT[d]	64						
RT+CT+HT	64	5.6 (7.4)	7.8 (6.2)	88 (.07)	0.14 (.05)	48-164	.57
RT + HT[e]	64	-1.1 (1.5)	2.8 (1.2)	14 (.11)	0.13 (.10)	8-25	.62

[a] converted from mean log DT, i.e. = $10^{(\text{mean log DT})}$

[b] standard deviation of $\log_{10}(DT)$

[c] goodness-of-fit

[d] the model failed to fit this group

[e] times beyond 15 months were censored to produce this model fit

(figures in parentheses are standard errors of estimates, and thus indicate the precision of the parameter estimates. The doubling time (μ_g) value is the standard error of $\log_{10}(DT)$)

Table 5.4

Parameter estimates for distant metastatic disease in the EORTC locally advanced breast cancer trial, assuming 10^9 metastatic cells are clinically detectable, and all these cells need to be eliminated to eradicate the distant metastatic disease.

Group	No. in group	log of remaining resistant tumour: mean (μ_v)	log of remaining resistant tumour: SD (σ_v)	doubling times: DT in days[a] (μ_g)	doubling times: SD[b] (σ_g)	95% range of DT	GOF[c] Test P-value
RT alone	72	1.6 (.77)	2.8 (1.0)	17 (.12)	0.39 (.09)	3-100	.62
RT + CT	69	1.9 (1.2)	3.8 (2.0)	30 (.14)	0.31 (.16)	7-124	.28
RT + HT[d]							
RT+CT+HT	64	2.8 (5.2)	1.6 (3.4)	47 (.53)	0.53 (.21)	4-504	.43
All 4 groups	269	1.5 (1.3)	2.1 (2.1)	24 (.16)	0.49 (.11)	3-218	.16

[a] converted from mean log DT, i.e. = $10^{(\text{mean log DT})}$

[b] standard deviation of $\log_{10}(DT)$

[c] goodness-of-fit

[d] the model failed to fit this group

(figures in parentheses are standard errors of estimates, and thus indicate the precision of the parameter estimates. The doubling time (μ_g) value is the standard error of $\log_{10}(DT)$)

Chapter 6

MULTIVARIATE REMISSION DURATION MODEL

Chapter 6

MULTIVARIATE REMISSION DURATION MODEL

6.1 Introduction

The remission duration model, as described in chapter 4, attempts to relate response durations to both sub-clinical residual (or more properly resistant) tumour, and to tumour growth rates. It seems likely that prognostic factors may affect either or both of these parameters. Some obvious examples would be presenting tumour size, which is very likely to be related to the volume of resistant tumour (see for instance, Goldie & Coldman (1979)), and putative markers of proliferation such as Ki67 (see chapter 3), which might be expected to correlate with tumour growth rates. It is therefore a fairly simple and logical extension of the model to allow putative prognostic factors to affect these parameters.

Prognostic factors have three main uses. Firstly, by defining risk groups, it is possible to give appropriate information to patients. Secondly, prognostic factors allow comparisons to be made between different centres (see for example, Wagstaff *et al* (1988)). Thirdly, it is sometimes possible to find high risk groups who need a new or better form of treatment, or low risk groups who may not need to be treated so intensively, or for so long. However, the definition of *risk* in itself does not indicate whether a patient will benefit from treatment (for example breast cancer

patients who are oestrogen receptor negative, may have a poorer prognosis than those who are oestrogen receptor positive, but they are less likely to benefit from endocrine therapy (McGuire *et al.*, 1982)). The principles of cancer therapy outlined in chapter 3 may, however, enable a good treatment strategy to be found for groups of patients at different degrees of risk. For example, chemotherapy may be most effective in tumours with a fast growth rate, or a novel treatment such as immunotherapy may be most effective in cases where the tumour burden is low. However, this information is not easily obtained, and the current statistical analysis techniques are of little help in this area.

The potential advantages of a multivariate version of the remission duration model are therefore considerable. Since the model provides information on two of the factors which are most important in deriving an appropriate strategy, namely resistance and growth rates, it should be possible to suggest suitable groups for a new treatment, and help in understanding the outcomes of existing treatments in particular prognostic subgroups.

A great deal of effort has been invested in discovering prognostic factors, particularly in the field of cancer (see for example the International Hodgkin's Disease Database Project (Somers *et al.*, 1990)). Although attempts have been made to target treatments to particular groups of patients, this has often merely entailed delineating groups of patients who perform badly, and giving these patients more intensive treatment. This is partly, as already suggested, because the current methods of prognostic factor analysis (mainly the proportional hazards model described by

Cox (1972)) only give significance levels and relative risks for the relevant factors. Thus there is little in the results themselves to indicate the appropriate treatment strategy. By relating prognostic factors to tumour growth rates and resistance, the multivariate methods based on the remission duration model presented in this thesis should be more useful in this area, and help in the choice of treatment strategies for prognostic subgroups, thus making good use of the vast amount of available prognostic information.

6.2 Mathematical description

The simplest method of applying prognostic factor analysis to the remission duration model is to allow factors to influence the mean resistant tumour and the mean growth rate in a linear fashion. Thus let the log of the resistant tumour be normally distributed, with log mean μ_v, where μ_v is a linear function of n prognostic variables, as follows:

$$\mu_v = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where $x_1,\ldots,x_n$ are the values of the n prognostic variables for a given individual, $\beta_1,\ldots,\beta_n$ are a set of regression coefficients, and β_0 is a baseline resistant tumour value (potentially for a patient having values of 0 for all the prognostic factors). Similarly let the log tumour doubling time, μ_g, also be a linear function of n prognostic variables:

$$\mu_g = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_n x_n$$

where $\gamma_1,\ldots,\gamma_n$ are a second set of regression coefficients, and γ_0 is a baseline log tumour doubling time (again potentially for a patient having values of 0 for all the prognostic factors).

Then the mathematics derived in chapter 4 apply, including the first and second partial derivatives, with $\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_nx_n$ replacing μ_v and $\gamma_0+\gamma_1x_1+\gamma_2x_2+\ldots+\gamma_nx_n$ replacing μ_g. The maximum likelihood routine needs to include the β's and γ's, and thus maximise on $2n+4$ parameters, rather than just 4. Therefore the first and second partial derivatives involving the βs and γs must be derived. For β_0 and γ_0 these are identical to the values derived for μ_v and μ_g in chapter 4. The first partial derivatives for the remaining βs and γs are as follows (derived from differentiating under the integral given in equations {4.5} and {4.6} from chapter 4):

$$\frac{\partial P}{\partial \beta_i} = \frac{\int_{V_o}^{V_r} x_i\{(v-\beta_0-\beta_1x_1-\beta_2x_2-\ldots-\beta_nx_n)/\sigma_v^2\}\, N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{P(CR)}$$

$$- \frac{\frac{\partial(P(CR))}{\partial \beta_i} \int_{V_o}^{V_r} N_v \int_{-\infty}^{U_t} N_g\, dg\, dv}{(P(CR))^2} \qquad \{6.1\}$$

where

$$N_v = (1/\sigma_v\sqrt{2\pi})\exp(-(v-\beta_0-\beta_1x_1-\beta_2x_2-\ldots-\beta_nx_n)^2/2\sigma_v^2) \qquad \{6.2\}$$

and

$$N_g = (1/\sigma_g\sqrt{2\pi})\exp(-(g-\gamma_0-\gamma_1x_1-\gamma_2x_2-\ldots-\gamma_nx_n)^2/2\sigma_g^2) \qquad \{6.3\}$$

and where, from equation {4.4} (with N_g and N_g modified as in equations {6.2} and {6.3}) differentiating under the integral,

$$\frac{\partial(P(CR))}{\partial\beta_i} = \int_{-\infty}^{V_r} x_i\{(v-\beta_0-\beta_1x_1-\beta_2x_2-\ldots-\beta_nx_n)/\sigma_v^2\}\ N_v\ dv \qquad \{6.4\}$$

The x_i can be taken outside the integral in equations {6.1} and {6.4} and the first derivative therefore reduces to:

$$\frac{\partial P}{\partial\beta_i} = x_i\ \frac{\partial P}{\partial\mu_v}$$

i.e. as in equation {4.8}, but with N_v and N_g defined as in equations {6.2} and {6.3}. Note generally that the form of all the first and second partial derivatives of the β's and γ's is identical to that of μ_v and μ_g respectively but with the additional multiplication by x_i, and that the x_i's can be taken outside the integrals. Thus the remaining partial derivatives are simply as follows:

$$\frac{\partial P'}{\partial\beta_i} = x_i\ \frac{\partial P'}{\partial\mu_v}$$

$$\frac{\partial P}{\partial\gamma_i} = x_i\ \frac{\partial P}{\partial\mu_g} \qquad \text{and} \qquad \frac{\partial P'}{\partial\gamma_i} = x_i\ \frac{\partial P'}{\partial\mu_g}$$

The second partial derivatives follow the same pattern:

$$\frac{\partial}{\partial\beta_i}\frac{\partial P}{\partial\theta_j} = x_i\ \frac{\partial}{\partial\mu_v}\frac{\partial P}{\partial\theta_j} \quad \text{and} \quad \frac{\partial}{\partial\beta_i}\frac{\partial P'}{\partial\theta_j} = x_i\ \frac{\partial}{\partial\mu_v}\frac{\partial P'}{\partial\theta_j} \qquad (j=1,4)$$

where $\theta_1=\mu_v$, $\theta_2=\sigma_v$, $\theta_3=\mu_g$, $\theta_4=\sigma_g$, as before.

and for the γ_i's

$$\frac{\partial}{\partial\gamma_i}\frac{\partial P}{\partial\theta_j} = x_i \frac{\partial}{\partial\mu_g}\frac{\partial P}{\partial\theta_j} \quad \text{and} \quad \frac{\partial}{\partial\gamma_i}\frac{\partial P'}{\partial\theta_j} = x_i \frac{\partial}{\partial\mu_g}\frac{\partial P'}{\partial\theta_j} \qquad (j=1,4)$$

The (β_i, γ_j), (β_i, β_j), (γ_i, γ_j) terms are:

$$\frac{\partial}{\partial\xi_i}\frac{\partial P}{\partial\xi_j} = x_i x_j \frac{\partial}{\partial\mu_v}\frac{\partial P}{\partial\mu_g} \quad \text{and} \quad \frac{\partial}{\partial\xi_i}\frac{\partial P'}{\partial\xi_j} = x_i x_j \frac{\partial}{\partial\mu_v}\frac{\partial P'}{\partial\mu_g} \qquad (\xi = \beta \text{ or } \gamma)$$

and the β_i^2 and γ_i^2 terms are:

$$\frac{\partial^2 P}{\partial\beta_i^2} = x_i^2 \frac{\partial^2 P}{\partial\mu_v^2} \quad \text{and} \quad \frac{\partial^2 P'}{\partial\beta_i^2} = x_i^2 \frac{\partial^2 P'}{\partial\mu_v^2}$$

$$\frac{\partial^2 P}{\partial\gamma_i^2} = x_i^2 \frac{\partial^2 P}{\partial\mu_g^2} \quad \text{and} \quad \frac{\partial^2 P'}{\partial\gamma_i^2} = x_i^2 \frac{\partial^2 P'}{\partial\mu_g^2}$$

Thus Newton's method (see chapter 4) can again be used to fit the model, and derive maximum likelihood estimates for the β's and γs.

6.3 Applications

6.3.1 Operable breast cancer

The first application is taken from a retrospective series of 966 patients aged less than 70 years with operable breast cancer. This cohort of patients, dating back to 1975, was selected to analyse the effect of menopausal status on patient survival in breast cancer, and is being reported separately. An application from this cohort was given in chapter 5.3 relating to the histological grade of the tumour. Four factors were considered in the multivariate model, namely the number of involved lymph

nodes, tumour size, histological grade, and the effect of the menopause (ultimately coded as those patients who were 1-5 years after the occurrence of the menopause versus all others). This study has the benefit, for modelling purposes of including prognostic factors almost certainly related to both growth rates and residual disease (see the application in chapter 5.3 for an explanation of why the model results in this case apply to residual rather than resistant disease and why this residual disease should be log-normally distributed). Tumour grade falls into the former category (see chapter 5.3). Tumour size certainly falls into the latter category, and the number of lymph nodes is probably also related to the volume of residual disease after surgery. It was not known whether any effect of the menopause would be related to tumour growth rate or to residual disease.

Note that for both the applications in this chapter, as for those given in chapter 5, it is assumed that 10^9 cells are clinically detectable, and all these cells need to be eliminated to eradicate the disease. The model estimates for doubling times under this assumption are not inconsistent with those reported in breast cancer (Shackney *et al.*, 1978).

The multivariate model was applied in a stepwise fashion, with each factor being considered for inclusion both for its effect on residual disease and for its effect on growth rates at each step. Thus initially the model was applied with just the four baseline parameters (β_0,σ_v,γ_0 and σ_g i.e. the model as in chapter 4 - $\mu_v=\beta_0$ and $\mu_g=\gamma_0$) giving a baseline likelihood. Then, for the first step, the model was applied for each factor in turn both for residual disease effect and growth rate effect,

estimating a single β_i or γ_i, and giving a series of new likelihoods. The differences from the baseline likelihood were doubled to give a statistic which is distributed as approximately chi-square with one degree of freedom (Silvey, 1970). This enables a p-value to be calculated for inclusion of the factor in the model. The factor giving the largest chi-square was then included, giving a new baseline likelihood, and the process repeated (see table 6.1 for a detailed worked example). Stepping was halted when no factors had significance levels less than 0.05. Factors were excluded from the model in a similar fashion if their significance levels were greater than .10.

It should be noted that if a factor's effect on doubling time was included at one step, that factor's effect on residual disease would still be considered in subsequent steps (and vice-versa). Thus factors may be related to both residual disease and growth rates.

The model results were compared with a similar analysis using the conventional multivariate approach, namely the proportional hazards model of Cox (1972). This produces a similar series of likelihoods and likelihood-ratio-test based chi-square statistics, based on a partial likelihood derived from the proportional hazards assumption (Cox, 1972).

The factors were coded as follows. The number of lymph nodes were categorised into 4 groups (coded as 0,1,2,3), the groups being 0, 1-3, 4-9 and $\geq$ 10 involved nodes. These groupings produce relapse-free curves that are roughly equally separated, and have in the past been commonly used to analyse the Guy's breast

cancer data (for examples, see Richards et al (1990), Badwe et al (1991)). Tumour size was coded as $\leq$ 2 cm and $>$ 2 cm, as this size appears to divide the curves with maximum separation (the log of the tumour size is ideally a better way to analyse its effect, but has the practical difficulty of how to code the tumours of size zero, and in any case gives nearly equivalent results to the method chosen). Histological grade was included as two variables, namely grade 1 versus grades 2 and 3, and grade 3 versus grades 1 and 2. This is necessary since the differences between the grades are not uniform (it would be inappropriate to enter the variable simply coded as 1, 2 or 3 since this would assume that the differences between grades 1 and 2 were the same as the differences between grades 2 and 3, which they clearly are not). Menopausal status was coded as 0 or 1, corresponding to the two groups described above.

The complete stepwise results for the application of the model are given in table 6.1, and the equivalent proportional hazards (Cox) model results are given in table 6.2.

Both histological grade variables exhibited a large and highly significant effect on doubling time, with quite large additional χ^2s compared to their possible effect on residual disease. In addition, after being entered into the model their subsequent effect on residual disease was not significant. This confirms the results from chapter 4.3 and provides a further validation of the model's assumptions. Furthermore, the significance of the histological grade variables was considerably greater than that derived from the Cox model, suggesting a better fit. The proportional hazards assumption is almost certainly violated by these variables, making the Cox model

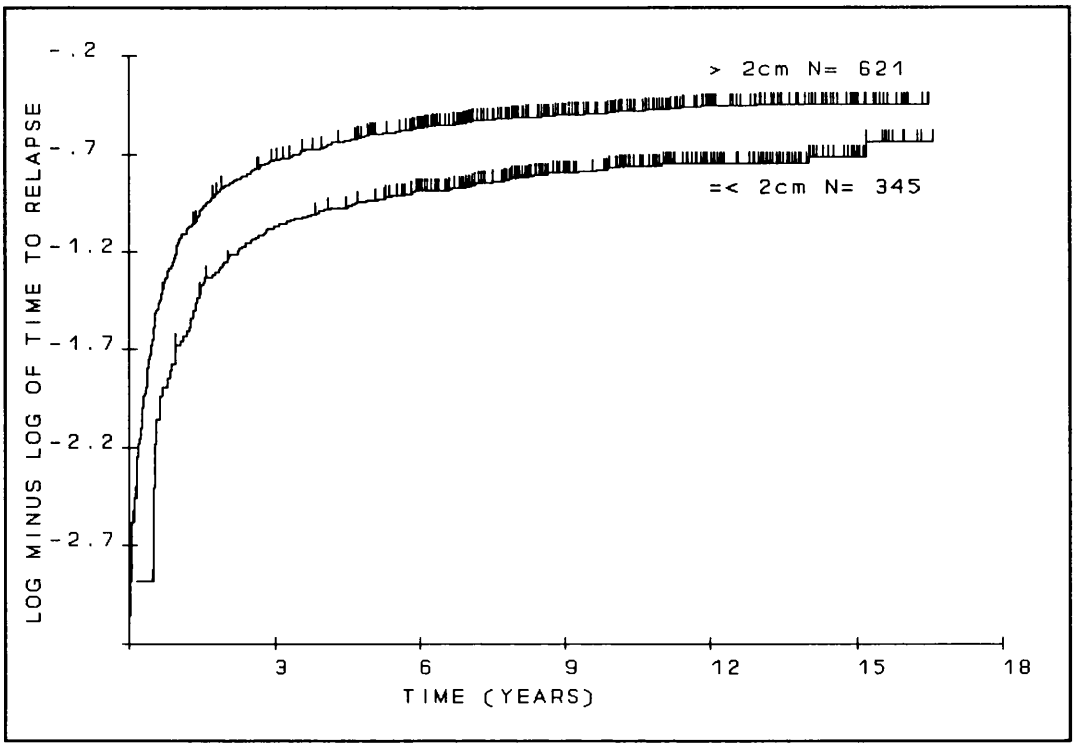


Figure 6.1 Log minus log plot for tumour size demonstrating approximately equal separation of the curves and thus proportionality of the hazards.

inappropriate. In contrast, the remission duration model incorporates such effects explicitly. Testing for goodness of fit of the proportional hazards model is not a straightforward task (Kalbfleisch & Prentice, 1980; Anderson, 1982). One approach is to plot log(-log(S(t,**z**))) where S(t,**z**) is the survivorship function for a set of covariate values, **z**. If the proportional hazards assumption is valid, the resulting plots should have a similar separation at all time points (Kalbfleisch & Prentice, 1980). An example where this is true is shown in figure 6.1, which plots tumour size above and below 2cm in this fashion. The log minus log plots for histological grade are shown in figure 6.2, and do not show a constant separation, the early parts of the curves being closer together than the later parts. This suggests that the hazards are not proportional, thus confirming that the proportional hazards model is inappropriate for this variable.

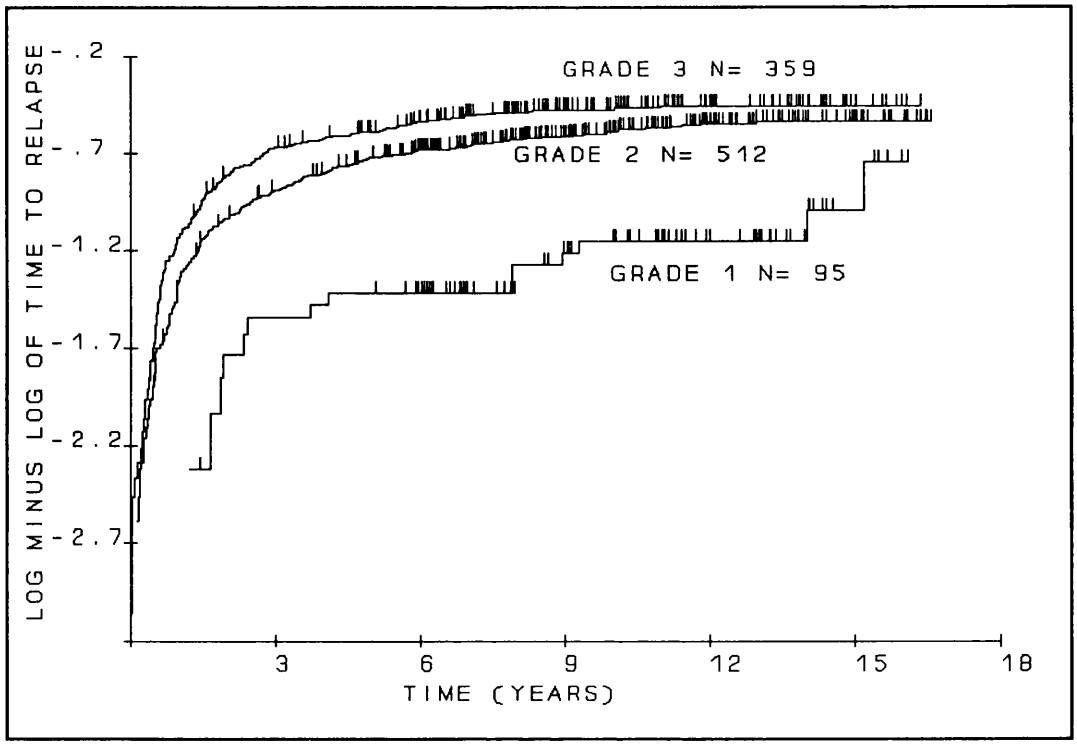


Figure 6.2 Log minus log plot for the histological grade groups, demonstrating the lack of proportionality of the hazards.

The effect of menstrual status is seen very clearly to be a residual disease effect. The contrast between its effect on residual disease compared to its effect on growth rates is the most clear of all the variables (in step 4, when it is entered into the model, the χ^2 for residual disease is 14.96, compared to only 5.5 for the growth rate effect). Following their findings of large differences in survival related to the timing of surgery within the menstrual cycle, the clinicians at Guy's hospital have been considering whether the poor outcome for patients who have recently become menopausal might result from an effect related to the timing of surgery effect, possibly hormonal in nature. It is important to know that the effect in this group is probably a residual disease effect, since this may have consequences for evaluating any new therapy given to these patients. It also suggests that an effective new treatment (simply putting the patients on a combination oestrogen/progesterone pill

has been suggested as a possible treatment) may well result in more patients being cured, not just in a prolongation of survival.

The overall fit of the remission duration model can be compared with that of the Cox model using the likelihood ratio test (Silvey, 1970). The overall χ^2s for improvement are 277.8 and 264.1 respectively for the two models. There are five extra parameters in the final step of the remission duration model compared with the final step of the Cox model, thus a χ^2 of 13.7 with 5 degrees of freedom is the appropriate test statistic, which yields a p-value of $<.02$. Thus the remission duration model produces a significantly better fit.

Before this analysis, it was thought by the clinicians involved in this study that the poor outcome for the menopausal patients might be a result of these patients having a worse than average mix of conventional prognostic factors (histological grade, nodes and tumour size). This is clearly seen, from the results of both multivariate models, not to be the explanation. However, the remission duration model results are more convincing in this respect since they incorporate the non-proportional-hazard effects of histological grade.

The consistency in the estimates of the four basic remission duration model parameters as additional variables are entered into the model is encouraging (see table 6.1). The SD estimates (σ_v and σ_g) fell slightly with the introduction of each additional variable, as might be expected since each additional variable explains some of the variability in the data, but were remarkably consistent. The mean

estimates (β_0 and γ_0) changed when a factor was entered which related to that mean, but were otherwise extremely consistent. These parameter values are listed separately for ease of comparison in table 6.3.

6.3.2 Locally advanced breast cancer

The second application is taken from the EORTC trial in locally advanced breast cancer reported in chapter 5.5. The comparison of radiotherapy (RT) with radiotherapy plus chemotherapy (RT+CT) for local recurrence suggested that the addition of CT produced an extra 3 logs of cell-kill on local disease. In this simple model comparison the other 3 parameters are also allowed to vary, and thus the estimate of 3 logs is not a direct comparison of the difference in mean log resistant disease (although it is *possible* that CT could influence the SD of the log of resistant disease and the SD of the log of doubling times, it seems nevertheless unlikely). Thus by putting treatment as a variable in the multivariate model, this possibility is eliminated and a more direct measure of cell-kill can be obtained. In addition a measure of any effect of CT on doubling time can be evaluated.

The multivariate results for this analysis are given in table 6.4. The treatment variable (i.e. the addition of CT) is seen to have an effect on resistant disease only, with the p-value for any additional effect on doubling time being only 0.39. The β_1 value is 3.53 thus quantifying the cell-kill effect at about 3½ logs, close to the value of 3 logs obtained in the original fits. The difference between the original fits and the new multivariate model fit can be compared using the likelihood ratio test (see

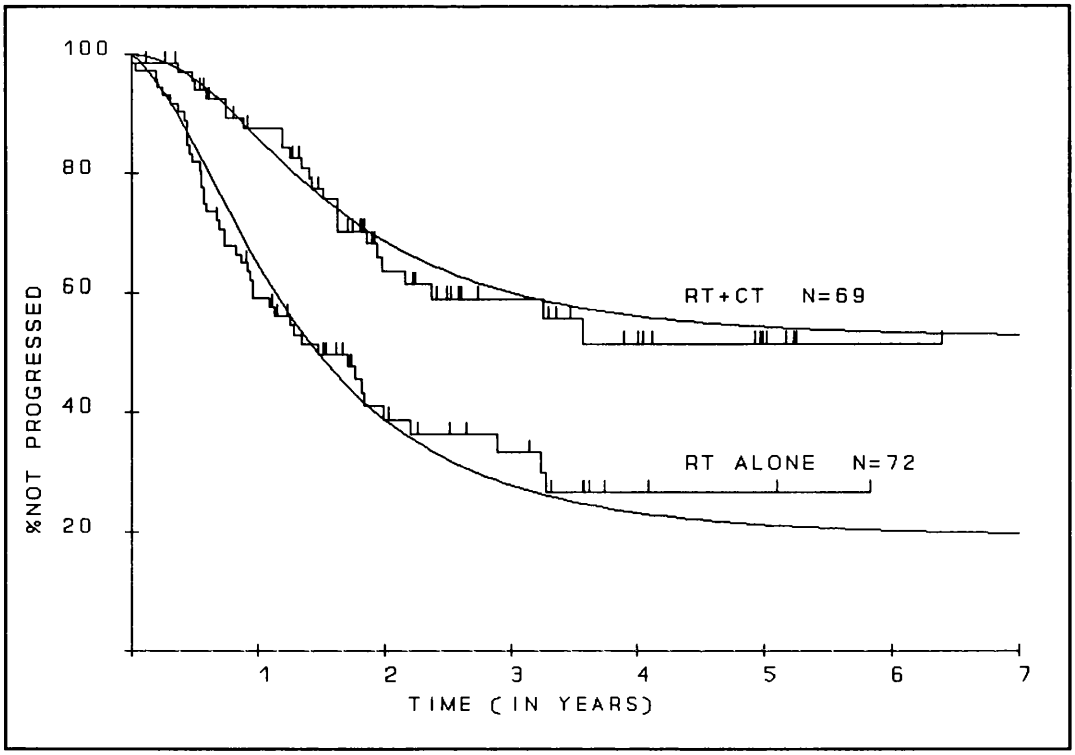


Figure 6.3 Multivariate model fits to the RT alone group and the RT+CT group for time to local progression.

chapter 4). The likelihoods for the two separate fits for RT and RT+CT were -340.36 and -204.36 respectively, giving an overall likelihood of -544.72. This compares with the combined likelihood for the multivariate model of -545.94, giving a χ^2 of 2.45, and a p-value of 0.12. Thus the fit for the multivariate model is not significantly worse than for the individual fits. The parameter values can be used to draw expected model curves for the two groups, and these are shown in figure 6.3. It can be seen that the model fits are still good. It is also encouraging that the estimated parameter values for the other three parameters are almost unchanged from their baseline estimates after the addition of the treatment variable into the model (see table 6.4). This confirms that these estimates are reliable and robust.

The remission duration model can again be compared with the proportional hazards model. The latter produced a χ^2 of 11.3 (p=.0008), which is again a worse fit,

although not significantly worse. There is some slight deviation away from proportional hazards in the early part of the curve, since the RT+CT curve starts with a short shoulder, whereas the RT alone curve falls almost immediately. This is however difficult to prove, given the difficulties in testing goodness-of-fit of the proportional hazards model already mentioned. It should be noted that in view of these restrictions in the proportional hazards model there are likely to be occasions where the remission duration model will detect differences that are not apparent with the proportional hazards model. The opposite is of course also possible for cases where the remission duration model provides a poor fit. Thus both models should have a role in the analysis of cancer trials data.

6.4 Discussion of the multivariate remission duration model results

The multivariate remission duration model has given consistent and comprehensible results for the two applications given. This is encouraging for its more widespread use in the analysis of response and remission duration data. The coefficients are easy to understand and interpret since they relate directly to measures of clinical efficacy. For instance from the estimated parameter values, the menopausal patients from the first example can be seen to have an additional residual disease volume of just over two logs compared to other patients (allowing for any imbalances in the other three factors). Although the doubling time coefficients are slightly more difficult to interpret, this is still a relatively easy task. A γ coefficient of 1 would imply a 1-log increase in doubling time, for example from 10 to 100 days. The values found in the

first example (see the final summary of coefficients table in table 6.1) are considerably less than this ranging from .08 to .61, which would correspond to increases in doubling time from, for example, 50 days to 68 days and 50 days to 544 days respectively.

As with the Cox proportional hazards model, estimated model curves can be easily constructed for any values of the parameters (see for example figure 6.3), providing a simple graphical representation of the nature and magnitude of the effects.

It is interesting to note that with as many as 6 variables included in the multivariate model, as in the operable breast cancer case, Newton's method still rapidly converged on the maximum likelihood. This method is relatively insensitive to the number of parameters to be estimated when the likelihood function has a clear maximum, as seems to be the case in the examples given.

In conclusion, this model should provide an additional and useful method of analysing multivariate response duration data, and will provide additional information to help in understanding the nature of the effects of different prognostic factors.

Table 6.1 Multivariate remission duration model results for the operable breast cancer application.

(The following abbreviations are used in this table to indicate whether the variables effect on mean residual disease or on mean log doubling time is being considered: RD = Residual disease DT = doubling time. In addition menopausal status is abbreviated MENOP S. The values given for γ_0 have been converted to doubling times for ease of comprehension; thus 10^{γ_0} is shown rather than γ_0 itself).

Descriptive statistics for variables (after recoding - see text)

VARIABLE NO.	NAME	MINIMUM	MAXIMUM	MEAN	STANDARD DEVIATION
1	NODES	0	3	.8271	.9658
2	GRADE 1v2&3	0	1	.9017	.2980
3	GRADE 1&2v3	0	1	.3716	.4835
4	TUMOUR SIZE	0	1	.6429	.4794
5	MENOP S.	0	1	.0963	.2951

Initial likelihood (4 basic parameters only in the model) = -3987.8591

Parameter values at maximum likelihood: β_0=0.26, σ_v=4.03, 10^{γ_0}=54.0, σ_g=0.46

STEP NUMBER 0 (NO TERMS IN THE MODEL)

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:NODES	186.03		<0.0001	-3894.8430
2	RD:GRADE 1v2&3	46.06		<0.0001	-3964.8308
3	RD:GRADE 1&2v3	20.40		<0.0001	-3977.6587
4	RD:TUMOUR SIZE	45.71		<0.0001	-3965.0034
5	RD:MENOP S.	18.23		<0.0001	-3978.7456
6	DT:NODES	149.75		<0.0001	-3912.9834
7	DT:GRADE 1v2&3	52.39		<0.0001	-3961.6650
8	DT:GRADE 1&2v3	23.21		<0.0001	-3976.2524
9	DT:TUMOUR SIZE	37.64		<0.0001	-3969.0410
10	DT:MENOP S.	6.83		0.0089	-3984.4426

Table continued on following page

STEP NUMBER 1 VARIABLE NUMBER 1 (RD:NODES) IS ENTERED

log likelihood = -3894.8430
Improvement in χ^2 (=2*(log(improv. in likelihood))) = 186.03 P < 0.0001

Parameter values at maximum likelihood: β_0=-1.83, σ_v=3.75, 10^{γ_0}=56.9, σ_g=0.41

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 RD:NODES	2.61

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

NO.	VARIABLE NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:NODES		186.03	<0.0001	-3987.8591
2	RD:GRADE 1v2&3	31.13		<0.0001	-3879.2803
3	RD:GRADE 1&2v3	17.08		<0.0001	-3886.3025
4	RD:TUMOUR SIZE	20.08		<0.0001	-3884.8015
5	RD:MENOP S.	18.88		<0.0001	-3885.4026
6	DT:NODES	7.09		0.0077	-3891.2974
7	DT:GRADE 1v2&3	37.88		<0.0001	-3875.9031
8	DT:GRADE 1&2v3	19.79		<0.0001	-3884.9460
9	DT:TUMOUR SIZE	17.26		<0.0001	-3886.2126
10	DT:MENOP S.	7.39		0.0066	-3891.1494

STEP NUMBER 2 VARIABLE NUMBER 7 (DT:GRADE 1v2&3) IS ENTERED

log likelihood = -3875.9031
Improvement in χ^2 (=2*(log(improv. in likelihood))) = 37.88 P < 0.0001

Parameter values at maximum likelihood: β_0=-1.53, σ_v=3.69, 10^{γ_0}=299, σ_g=0.41

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 RD:NODES	2.52
7 DT:GRADE 1v2&3	-0.74

Table continued on following page

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:NODES		171.52	<0.0001	-3961.6650
2	RD:GRADE 1v2&3	0.39		0.5333	-3875.7090
3	RD:GRADE 1&2v3	8.39		0.0038	-3871.7102
4	RD:TUMOUR SIZE	19.49		<0.0001	-3866.1560
5	RD:MENOP S.	17.74		<0.0001	-3867.0327
6	DT:NODES	5.82		0.0158	-3872.9917
7	DT:GRADE 1v2&3		37.88	<0.0001	-3894.8430
8	DT:GRADE 1&2v3	12.72		0.0004	-3869.5417
9	DT:TUMOUR SIZE	14.50		0.0001	-3868.6553
10	DT:MENOP S.	6.24		0.0125	-3872.7827

STEP NUMBER 3 VARIABLE NUMBER 4 (RD:TUMOUR SIZE) IS ENTERED

log likelihood = -3866.1560

Improvement in χ^2 (=2*(log(improv. in likelihood))) = 19.49 P < 0.0001

Parameter values at maximum likelihood: β_0=-2.41, σ_v=3.59, 10^{γ_0}=291, σ_g=0.41

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 RD:NODES	2.31
7 DT:GRADE 1v2&3	-0.73
4 RD:TUMOUR SIZE	1.61

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:NODES		148.24	<0.0001	-3940.2737
2	RD:GRADE 1v2&3	2.12		0.1458	-3865.0984
3	RD:GRADE 1&2v3	7.93		0.0049	-3862.1912
4	RD:TUMOUR SIZE		19.49	<0.0001	-3875.9028
5	RD:MENOP S.	14.96		0.0001	-3858.6777
6	DT:NODES	6.21		0.0127	-3863.0508
7	DT:GRADE 1v2&3		37.29	<0.0001	-3884.8015
8	DT:GRADE 1&2v3	13.11		0.0003	-3859.5994
9	DT:TUMOUR SIZE	1.49		0.2221	-3865.4106
10	DT:MENOP S.	5.50		0.0190	-3863.4043

Table continued on following page

STEP NUMBER 4 VARIABLE NUMBER 5 (RD:MENOP S.) IS ENTERED

log likelihood = -3858.6777
Improvement in χ^2 (=2*(log(improv. in likelihood))) = 14.96 P = 0.0001

Parameter values at maximum likelihood: β_0=-2.57, σ_v=3.58, 10^{γ_0}=282, σ_g=0.40

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 RD:NODES	2.32
7 DT:GRADE 1v2&3	-0.71
4 RD:TUMOUR SIZE	1.49
5 RD:MENOP S.	2.26

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:NODES		150.07	<0.0001	-3933.7129
2	RD:GRADE 1v2&3	1.72		0.1892	-3857.8157
3	RD:GRADE 1&2v3	6.84		0.0089	-3855.2585
4	RD:TUMOUR SIZE		16.71	<0.0001	-3867.0327
5	RD:MENOP S.		14.96	0.0001	-3866.1560
6	DT:NODES	5.74		0.0166	-3855.8064
7	DT:GRADE 1v2&3		36.19	<0.0001	-3876.7705
8	DT:GRADE 1&2v3	13.22		0.0003	-3852.0693
9	DT:TUMOUR SIZE	1.37		0.2422	-3857.9939
10	DT:MENOP S.	0.03		0.8554	-3858.6611

STEP NUMBER 5 VARIABLE NUMBER 8 (DT:GRADE 1&2v3) IS ENTERED

log likelihood = -3852.0693
Improvement in χ^2 (=2*(log(improv. in likelihood))) = 13.22 P = 0.0003

Parameter values at maximum likelihood: β_0=-2.55, σ_v=3.59, 10^{γ_0}=281, σ_g=0.39

Table continued on following page

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 RD:NODES	-2.32
7 DT:GRADE 1v2&3	-0.62
4 RD:TUMOUR SIZE	1.51
5 RD:MENOP S.	2.27
8 DT:GRADE 1&2v3	-0.19

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:NODES		149.29	<0.0001	-3926.7119
2	RD:GRADE 1v2&3	1.96		0.1616	-3851.0898
3	RD:GRADE 1&2v3	0.39		0.5317	-3851.8738
4	RD:TUMOUR SIZE		17.05	<0.0001	-3860.5928
5	RD:MENOP S.		15.06	0.0001	-3859.5994
6	DT:NODES	6.24		0.0125	-3848.9480
7	DT:GRADE 1v2&3		29.18	<0.0001	-3866.6577
8	DT:GRADE 1&2v3		13.22	0.0003	-3858.6777
9	DT:TUMOUR SIZE	1.31		0.2515	-3851.4119
10	DT:MENOP S.	0.06		0.8041	-3852.0386

STEP NUMBER 6 VARIABLE NUMBER 6 (DT:NODES) IS ENTERED

log likelihood = -3848.9480

Improvement in χ^2 (=2*(log(improv. in likelihood))) = 6.24 P = 0.0125

Parameter values at maximum likelihood: β_0=-2.14, σ_v=3.46, 10^{γ_0}=320, σ_g=0.40

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 RD:NODES	1.78
7 DT:GRADE 1v2&3	-0.61
4 RD:TUMOUR SIZE	1.46
5 RD:MENOP S.	2.13
8 DT:GRADE 1&2v3	-0.19
6 DT:NODES	-0.08

Table continued on following page

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:NODES		40.29	<0.0001	-3869.0935
2	RD:GRADE 1v2&3	2.82		0.0931	-3847.5381
3	RD:GRADE 1&2v3	0.23		0.6312	-3848.8328
4	RD:TUMOUR SIZE		17.22	<0.0001	-3857.5562
5	RD:MENOP S.		14.45	0.0001	-3856.1716
6	DT:NODES		6.24	0.0125	-3852.0693
7	DT:GRADE 1v2&3		27.96	<0.0001	-3862.9302
8	DT:GRADE 1&2v3		13.72	0.0002	-3855.8064
9	DT:TUMOUR SIZE	0.60		0.4372	-3848.6462
10	DT:MENOP S.	0.20		0.6565	-3848.8491

No term passes the remove or enter levels (0.1000 0.0500).
Stepwise procedure complete.

SUMMARY OF STEPWISE PROCEDURE

STEP	VARIABLE	ENTERED/ REMOVED	CHI-SQ. FOR INCLUSION/ EXCLUSION	P-VALUE	LOG LIKELIHOOD
0					-3987.8591
1	RD:NODES	ENTERED	186.03	<0.0001	-3894.8430
2	DT:GRADE 1v2&3	ENTERED	37.88	<0.0001	-3875.9031
3	RD:TUMOUR SIZE	ENTERED	19.49	<0.0001	-3866.1560
4	RD:MENOP S.	ENTERED	14.96	0.0001	-3858.6777
5	DT:GRADE 1&2v3	ENTERED	13.22	0.0003	-3852.0693
6	DT:NODES	ENTERED	6.24	0.0125	-3848.9480

Table 6.2 Multivariate proportional hazards model results for the first application.

(Menstrual status is abbreviated MENOP S. in this table. Note also that the proportional hazards model is based on *partial* likelihoods, but for simplicity in this table this word has been dropped and they will simply be referred to as likelihoods).

For a table of descriptive statistics for the 5 variables included in the model see table 6.1.

Initial likelihood (no terms in the model) = -2758.9720

STEP NUMBER 0 (NO TERMS IN THE MODEL)

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	NODES	183.79		<.0001	-2667.0780
2	GRADE 1v2&3	45.80		<.0001	-2736.0730
3	GRADE 1&2v3	22.08		<.0001	-2747.9330
4	TUMOUR SIZE	43.84		<.0001	-2737.0510
5	MENOP S.	20.14		<.0001	-2748.9020

STEP NUMBER 1 VARIABLE NUMBER 1 (NODES) IS ENTERED

log likelihood = -2667.0780

Improvement in χ^2 (=2*(log(improv. in likelihood))) = 183.79 P < 0.0001

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 NODES	0.658

Table continued on following page

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	NODES		183.79	<.0001	-2758.9720
2	GRADE 1v2&3	34.43		<.0001	-2649.8630
3	GRADE 1&2v3	15.89		.0001	-2659.1320
4	TUMOUR SIZE	23.05		<.0001	-2655.5510
5	MENOP S.	22.99		<.0001	-2655.5810

STEP NUMBER 2 VARIABLE NUMBER 2 (GRADE 1v2&3) IS ENTERED

log likelihood = -2649.8630

Improvement in χ^2 (=2*(log(improv. in likelihood))) = 34.43 P < 0.0001

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 NODES	0.638
2 GRADE 1v2&3	1.294

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	NODES		172.42	<.0001	-2736.0730
2	GRADE 1v2&3		34.43	<.0001	-2667.0780
3	GRADE 1&2v3	7.52		.0061	-2646.1010
4	TUMOUR SIZE	21.67		<.0001	-2639.0290
5	MENOP S.	20.43		<.0001	-2639.6470

Table continued on following page

STEP NUMBER 3 VARIABLE NUMBER 4 (TUMOUR SIZE) IS ENTERED

log likelihood = -2639.0290
Improvement in χ^2 (=2*(log(improv. in likelihood))) = 21.67 P < 0.0001

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 NODES	0.610
2 GRADE 1v2&3	1.272
4 TUMOUR SIZE	0.508

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	NODES		154.00	<.0001	-2716.0310
2	GRADE 1v2&3		33.04	<.0001	-2655.5510
3	GRADE 1&2v3	7.68		.0056	-2635.1900
4	TUMOUR SIZE		21.67	<.0001	-2649.8630
5	MENOP S.	17.64		<.0001	-2630.2080

STEP NUMBER 4 VARIABLE NUMBER 5 (MENOP S.) IS ENTERED

log likelihood = -2630.2080
Improvement in χ^2 (=2*(log(improv. in likelihood))) = 17.64 P < 0.0001

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 NODES	0.616
2 GRADE 1v2&3	1.235
4 TUMOUR SIZE	0.477
5 MENOP S.	0.635

Table continued on following page

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	NODES		155.87	<.0001	-2708.1450
2	GRADE 1v2&3		30.62	<.0001	-2645.5160
3	GRADE 1&2v3	6.58		.0103	-2626.9180
4	TUMOUR SIZE		18.88	<.0001	-2639.6470
5	MENOP S.		17.64	<.0001	-2639.0290

STEP NUMBER 5 VARIABLE NUMBER 3 (GRADE 1&2v3) IS ENTERED

log likelihood = -2626.9180

Improvement in χ^2 (=2*(log(improv. in likelihood))) = 6.58 P = 0.0103

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 NODES	0.613
2 GRADE 1v2&3	1.129
4 TUMOUR SIZE	0.478
5 MENOP S.	0.615
3 GRADE 1&2v3	0.257

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	NODES		155.88	<.0001	-2704.8560
2	GRADE 1v2&3		23.39	<.0001	-2638.6120
3	GRADE 1&2v3		6.58	.0103	-2630.2080
4	TUMOUR SIZE		18.96	<.0001	-2636.3970
5	MENOP S.		16.54	<.0001	-2635.1900

All variables entered into the model. Stepwise procedure complete.

Table continued on following page

SUMMARY OF STEPWISE PROCEDURE

STEP	VARIABLE	ENTERED/ REMOVED	CHI-SQ. FOR INCLUSION/ EXCLUSION	P-VALUE	LOG LIKELIHOOD
0					-2758.9720
1	NODES	ENTERED	183.79	<0.0001	-2667.0780
2	GRADE 1v2&3	ENTERED	34.43	<0.0001	-2649.8630
3	TUMOUR SIZE	ENTERED	21.67	<0.0001	-2639.0290
4	MENOP S.	ENTERED	17.64	0.0001	-2630.2080
5	GRADE 1&2v3	ENTERED	6.58	0.0103	-2626.9180

Table 6.3 Values for the four remission duration model basic parameters at each step of the first multivariate model application.

STEP	β_0	σ_v	10^{γ_0}	σ_g
0	0.26	4.03	54.0	0.46
1	-1.83	3.75	56.9	0.41
2	-1.53	3.69	298.9	0.41
3	-2.41	3.59	290.9	0.41
4	-2.57	3.58	281.9	0.40
5	-2.55	3.59	281.1	0.39
6	-2.14	3.46	320.1	0.40

Table 6.4 Multivariate remission duration model results for the locally advanced breast cancer application.

Descriptive statistics for variable (coded as 0 for RT and 1 for RT+CT)

VARIABLE NO.	NAME	MINIMUM	MAXIMUM	MEAN	STANDARD DEVIATION
1	RTvRT+CT	0	1	.4894	.5017

Initial likelihood (4 basic parameters only in the model) = -553.4655

Parameter values at maximum likelihood: β_0=1.60, σ_v=4.06, 10^{γ_0}=26.5, σ_g=0.28

STEP NUMBER 0 (NO TERMS IN THE MODEL)

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:RTvRT+CT	15.05		0.0001	-545.9411
2	DT:RTvRT+CT	12.27		0.0005	-547.3318

STEP NUMBER 1 VARIABLE NUMBER 1 (RD:RTvRT+CT) IS ENTERED

log likelihood = -545.9411

Improvement in χ^2 (=2*(log(improv. in likelihood))) = 15.05 P = 0.0001

Parameter values at maximum likelihood: β_0=3.35, σ_v=3.66, 10^{γ_0}=27.3, σ_g=0.27

SUMMARY OF VARIABLES CURRENTLY ENTERED

VARIABLE	COEFF.
1 RD:RTvRT+CT	-3.53

Table continued on following page

STATISTICS TO ENTER OR REMOVE VARIABLES ARE AS FOLLOWS:

VARIABLE NO.	NAME	APPROX CHI-SQ. ENTER	APPROX CHI-SQ. REMOVE	P-VALUE	LOG LIKELIHOOD
1	RD:RTvRT+CT		15.05	0.0001	-553.4655
2	DT:RTvRT+CT	0.73		0.3934	-545.5770

No term passes the remove or enter levels (0.1000 0.0500).
Stepwise procedure complete.

SUMMARY OF STEPWISE PROCEDURE

STEP	VARIABLE	ENTERED/ REMOVED	CHI-SQ. FOR INCLUSION/ EXCLUSION	P-VALUE	LOG LIKELIHOOD
0					-553.4655
1	RD:RTvRT+CT	ENTERED	15.05	0.0001	-545.9411

Chapter 7

RESISTANCE MODEL FOR INDIVIDUAL PATIENTS

Chapter 7

RESISTANCE MODEL FOR INDIVIDUAL PATIENTS

7.1 Introduction

The previous three chapters have presented a population based model designed to make inferences about resistance and growth rates based on the relapse patterns for populations of patients. Such a population-based model should aid in the design of future trials and in the interpretation of current and previous trials, but is less likely to help in the treatment of individual patients. A complementary model has been developed to try to obtain information about resistance and growth rates for individual patients, using some measure of the tumour volume.

The models of Goldie and Coldman (1979; 1982) and Skipper (Skipper & Perry, 1970) espouse general principles, such as the alternating of non-cross-resistant drug combinations. A theoretical model was also developed for individual patients (Birkhead & Gregory, 1984; Birkhead *et al.*, 1986). The consequences of particular assumptions about the values of this model's parameters could be explored, but the model could not be fitted to real data. A very restricted fitting method was employed when applying the model to small cell lung cancer (SCLC) data from a University College Hospital trial of high dose therapy, in collaboration with Professor Souhami (Gregory *et al.*, 1988). This was only applicable where just three tumour volumes were measured, and involved estimating the doubling time independently from other

published reports. This chapter describes an extension of this model to enable the fitting of real tumour volume data, with all its inherent variability. Such a model has the potential for suggesting times at which to change or abandon treatment for individual patients, as well as providing estimates for resistance and tumour-kill which may provide additional useful outcome measures for clinical trials, and help in the design of future studies. It is also possible that, once validated, such a model might enable results to be achieved on smaller numbers of patients, since the additional interpretative information ought to improve the power of any tests used.

7.2 Basic model description and assumptions

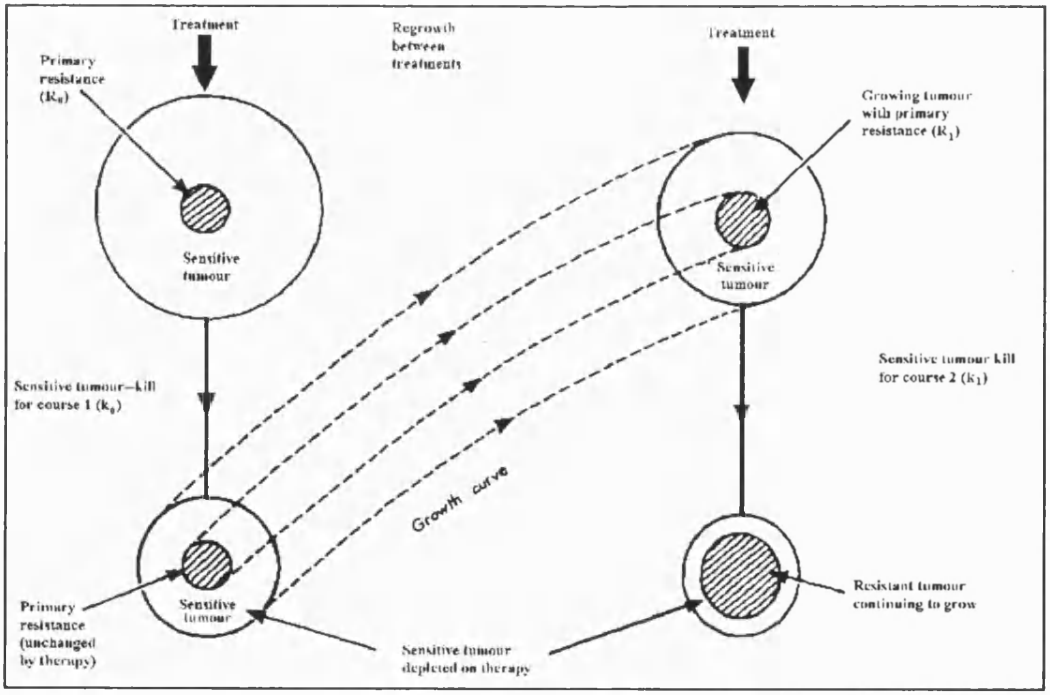


Figure 7.1 Diagrammatic representation of the model

The model seeks to relate changing tumour volumes to proportions of sensitive and resistant tumour, and to tumour growth rate. This is represented diagrammatically in Figure 7.1. Resistant tumour is assumed to be tumour which can never be killed with the given drug dose due to inherent (cellular) resistance. The remaining tumour is considered sensitive, although it is assumed that only a proportion of it will be killed by a single administration of the drug. This is based on the fractional cell-kill hypotheses of Skipper and colleagues (Skipper *et al.*, 1964; 1967) discussed in chapter 3. Various reasons for cells not being killed can be hypothesised, for example cells not being in cycle, uneven drug distribution, problems of blood supply, and the likely stochastic nature of cell-killing by cytotoxic agents. The tumour growth rate is empirically assumed to be exponential for the period of therapy, and it is assumed that throughout the treatment period the mutation rate of cells from sensitivity to resistance or vice versa is negligible in comparison to the other effects (i.e. the two populations of sensitive and resistant cells are independent: cells do not change from being sensitive to being resistant or vice versa).

The proportion of sensitive tumour killed by each cycle of the treatment is thus assumed to be the same (Skipper & Perry, 1970), and is represented by k. The proportion of tumour initially resistant is represented by R_0. The tumour doubling time is denoted by d. The model predicts, for particular values of these three independent variables, k, R_0 and d, given the above assumptions, the sequence of tumour volumes before each treatment cycle (Birkhead & Gregory, 1984).

An assumption must also be made about the distribution of errors in volume measurements, in order to fit the model to real data. Supposing k, R_0 and d were known, some slight differences would still be expected between the model predictions and the actual tumour volumes, due to inaccuracies in measurement. For example when measuring disease volume in the lung, variations in marking out the area of tumour, or delineating the tumour from other structures, as well as collapse and consolidation of lung tissue around the tumour could all contribute to this error. Following discussion with the clinicians involved, two likely assumptions were suggested for this distribution of errors in measurement, for use in two different circumstances. For some tumours, like breast cancer, where clinical volume measurements were being made, it was suggested that errors in volume measurements were likely to be normally distributed. For other tumours, like SCLC, where added accuracy was obtained by use of instruments such as computerised tomography (CT) scans, it was thought that a log-normal distribution of errors would be more appropriate, especially for small tumours, where the CT scanner would be more sensitive (this assumption is equivalent to the assumption that the same percentage error can be expected at each volume). Models have therefore been constructed for both assumptions. The application to SCLC described in this chapter used CT scans to measure tumour volumes and therefore uses the log-normal assumption. An application currently underway in breast cancer, where clinical volume measurements are being employed, uses the straight normality assumption.

The relevant mathematics (incorporating both error distribution assumptions) will now be described.

7.3 Mathematical description of the model

The model predicts that sequential tumour volumes before treatment (X_0, X_1, X_2, ..., X_n) will be described by the equation:

$$X_i = \frac{1-a-(1-a^i)k_0}{1-a} X_0 \exp(\alpha t_i) \qquad (i = 1, 2, \ldots, n) \qquad \{7.1\}$$

where $a = (1-k)$, $k_0 = k(1-R_0)$, k is the proportion of the sensitive tumour killed with each course of therapy, R_0 is the proportion of the tumour initially resistant, α is the (exponential growth rate), t_i is the time between first treatment and treatment cycle $i+1$, and i is the treatment cycle number itself.

From equation {7.1}

$$\log X_i = \log \left[\frac{1-a-(1-a^i)k_0}{1-a} \right] + \log X_0 + \alpha t_i \qquad \{7.2\}$$

The preceding theory is that described in Birkhead and Gregory (1984). The new developments to enable the model to be fitted to real clinical data will now be described. Firstly some assumption must be made about errors in the tumour volume measurements.

Let the actual tumour volumes be V_0, V_1, ..., V_n.

Under the assumption of a log normal distribution of errors in measurement, the likelihood, L, of the (log of) the volumes under the model is:

$$L(\log V_0, \log V_1, \ldots\ldots, \log V_n) = N(\log V_0, \log X_0, \sigma).$$

$$N(\log V_1, \log X_1, \sigma) \ldots. N(\log V_n, \log X_n, \sigma)$$

$$= \prod_{i=0}^{n} N(\log V_i, \log X_i, \sigma)$$

where $N(x, u, \sigma)$ is the value of a normal distribution with mean u and variance σ at x.

Hence

$$\log L = \sum_{i=0}^{n} \log N(\log V_i, \log X_i, \sigma)$$

Now $N(x, u, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(u-x)^2}{2\sigma^2}\right]$

So $\log L = \sum_{i=0}^{n} \log \left[\frac{1}{\sigma\sqrt{2\pi}} \exp \left[\frac{-(\log X_i - \log V_i)^2}{2\sigma^2} \right] \right]$ {7.3}

which simplifies to

$$\log L = \sum_{i=0}^{n} \log \frac{1}{\sigma\sqrt{2\pi}} - \sum_{i=0}^{n} \frac{-(\log X_i - \log V_i)^2}{2\sigma^2}$$

The maximum likelihood estimates (MLEs) for $\log X_0$, k, R_0, α, and σ (i.e. the values of these parameters which produce the closest fit between the model's predictions and the data) can then be obtained by maximising log L from {7.3}. This can be achieved by differentiating log L with respect to each of the parameters

$\log X_0$, k, R_0, α, and σ and maximising log L based on the values of these derivatives using a semi-Newton algorithm in a similar fashion as described in chapter 4 for the Newton algorithm on the population based model (the semi-Newton algorithm uses the first derivatives to approximate the matrix of second partial derivatives, as described by Beale (1988)). The first derivatives are as follows:

$$\frac{\partial \text{Log L}}{\partial k} = \sum_{i=0}^{n} \left[\frac{(\log X_i - \log V_i)}{\sigma^2} \frac{k_0 \, i \, a^{i-1}}{\{1-a-(1-a^i)k_0\}} \right]$$

$$\frac{\partial \text{Log L}}{\partial R_0} = \sum_{i=0}^{n} \left[\frac{-(\log X_i - \log V_i)}{\sigma^2} \frac{(1-a^i)(1-a)}{\{1-a-(1-a^i)k_0\}} \right]$$

$$\frac{\partial \text{Log L}}{\partial \alpha} = \sum_{i=0}^{n} \left[\frac{-(\log X_i - \log V_i)}{\sigma^2} t_i \right]$$

$$\frac{\partial \text{Log L}}{\partial \log X_0} = \sum_{i=0}^{n} \left[\frac{-(\log X_i - \log V_i)}{\sigma^2} \right]$$

$$\frac{\partial \text{Log L}}{\partial \sigma} = \frac{-(n+1)}{\sigma} + \sum_{i=0}^{n} \left[\frac{(\log X_i - \log V_i)^2}{\sigma^3} \right]$$

Under the *alternative* assumption that errors of the same magnitude occur at any volume, the likelihood, L, of the (log of) the volumes under the model is

$$L(V_0, V_1, \ldots\ldots, V_n) = N(V_0, X_0, \sigma).N(V_1, X_1, \sigma) \ldots. N(V_n, X_n, \sigma)$$

$$= \prod_{i=0}^{n} N(V_i, X_i, \sigma)$$

where, as before, N(x, u, σ) is the value of a normal distribution with mean u and variance σ at x.

Hence

$$\log L = \sum_{i=0}^{n} \log N(V_i, X_i, \sigma)$$

$$\text{So } \log L = \sum_{i=0}^{n} \log \left[\frac{1}{\sigma\sqrt{2\pi}} \exp \left[\frac{-(X_i - V_i)^2}{2\sigma^2} \right] \right] \qquad \{7.4\}$$

which simplifies to

$$\log L = \frac{(n+1)}{\sigma\sqrt{2\pi}} - \sum_{i=0}^{n} \frac{-(X_i - V_i)^2}{2\sigma^2}$$

The maximum MLEs for X_0, k, R_0, α, and σ can again be obtained by maximising log L from {7.4}, using the first derivatives of log L with respect to each of the parameters X_0, k, R_0, α, and σ and maximising log L using the semi-Newton algorithm. The first derivatives in this case are as follows:

$$\frac{\partial \text{Log } L}{\partial k} = \sum_{i=0}^{n} \left[\frac{(X_i - V_i)}{\sigma^2} \frac{k_0 \, i \, a^{i-1}}{(1-a)} X_0 \exp(\alpha t_i) \right]$$

$$\frac{\partial \text{Log } L}{\partial R_0} = \sum_{i=0}^{n} \left[\frac{-(X_i - V_i)}{\sigma^2} (1-a^i) X_0 \exp(\alpha t_i) \right]$$

$$\frac{\partial \text{Log } L}{\partial \alpha} = \sum_{i=0}^{n} \left[\frac{-(X_i - V_i)}{\sigma^2} \frac{\{1-a-k_0(1-a^i)\}}{(1-a)} X_0 t_i \exp(\alpha t_i) \right]$$

$$\frac{\partial \text{Log } L}{\partial X_0} = \sum_{i=0}^{n} \left[\frac{-(X_i - V_i)}{\sigma^2} \frac{\{1-a-k_0(1-a^i)\}}{(1-a)} \exp(\alpha t_i) \right]$$

$$\frac{\partial \text{Log } L}{\partial \sigma} = \frac{-(n+1)}{\sigma} + \sum_{i=0}^{n} \left[\frac{(X_i - V_i)^2}{\sigma^3} \right]$$

7.4 Application Methodology

An attempt is made to validate this model, assuming a log-normal distribution of errors in tumour volume measurements, on patients with small cell lung cancer. The model requires an accurate method of measuring tumour volumes to minimise these errors, and CT scans of the chest have been employed to this end. This method gives an accurate measurement of the tumour volume, although errors of some magnitude will obviously still be made. Having estimated the resistance to, and efficacy of chemotherapy, and the tumour growth rate, the model predicts the sequence of tumour volumes before each course of chemotherapy. The validity and accuracy of these predictions were tested on a series of up to 7 scans on each of 9 patients with SCLC.

7.4.1 Theoretical considerations

In order to estimate the model's parameters, i.e. the proportion of sensitive tumour killed with each cycle of therapy, k, the resistance at presentation R_0, and the tumour

doubling time d, at least four tumour volumes are required (under either error assumption).

With exactly four volumes, all of these volumes will be needed to estimate the parameters. If the model is not a reasonable representation of the actual disease processes, it may be expected that no values of the parameters would be capable of predicting the observed volumes. For instance, if the percentage tumour reduction on the first cycle of treatment was less than that seen on the second cycle (assuming a similar interval between cycles) the model would be invalid. With more than four tumour volumes the accuracy of the model can be evaluated, assuming the model fits at all, as just explained, since its consistency in predicting the sequential tumour volumes can be examined. In such cases all the tumour volumes can be used to estimate the model's parameters. The model can then be validated by a chi-square test comparing the observed tumour volumes with those expected under the model assumptions. Furthermore, in the patients with more than four volumes, since k, R_0 and d can be estimated from just four volumes, these estimates can be used to predict the remaining volumes, providing a further substantive test of the model's validity.

The standard deviation (SD) of the log-normal distribution of errors will reflect differences between the observed tumour volumes and the model's predictions, and will thus provide a measure of the accuracy of the model. Where the SD is small, the predictions and actuals will match closely. Where the SD is large, one or more predicted volumes will show large differences from the corresponding measured

value. This can be seen in the application which follows in section 7.4.2; for instance when comparing patient number 2 with patient number 8 (see table 7.3 for the SDs of errors about model predictions, and figures 7.2 and 7.3 for a graphical representation).

A computer program, written in Microsoft FORTRAN 77 for IBM compatible microcomputers, has been written to produce the estimates. The estimation procedure takes only a few seconds to run. A worked example is provided in table 7.4.

7.4.2 Application to Small Cell Lung Cancer

A. **Patients**

Nine patients with SCLC had tumour volumes measured. They were taken from two separate trials, one comparing Etoposide (VP16) and Doxorubicin (Adriamycin) (VA) with Oncovin (Vincristine), Etoposide and Adriamycin (OVA) in limited disease patients, the other comparing two different schedules of Etoposide given as a single agent in extensive disease (Slevin *et al.*, 1989) the same dose of Etoposide being given as a continuous infusion for 1 day, or as separate 2 hour continuous infusions over 5 days.

The observed tumour volumes along with the times (in days) since the start of treatment, at which the scans were taken, are given in table 1, and shown

diagrammatically in figures 7.2 and 7.3. Patients with peripheral masses on chest X-ray were chosen for the study since it was possible to separate tumour from mediastinal structures on the scans in these patients. Tumour volume measurements were made by an experienced CT radiologist, Dr. Rodney Reznek. The patients were scanned on a GE 9800 Whole Body Scanner. Scans were performed at 1cm intervals throughout the region of the tumour. Where necessary, a bolus of intravenous contrast medium was administered to delineate vascular structures. The area of the lesion was then calculated on each image using a tracing device. As the scan thickness was 1cm in each image the volume could be easily estimated. Care was taken to avoid measuring areas of lung consolidation or collapse, though this was not always possible. Where such discrimination was difficult in a series of scans a special effort was made to measure the same structures on each scan in the series. However the initial measurements in this series were often made as the scans became available, several weeks apart.

One patient died during therapy and consequently has only three tumour volumes recorded; the rest have at least four, generally five, and in one case seven tumour volumes measured.

Because of the requirement for having at least four volumes in order to estimate the model's parameters, the model could only be applied to eight of the nine patients.

B. **Results**

Reproducibility of volume estimates

To test the reproducibility and accuracy of the CT volume estimates, four of the nine patients' volumes were independently re-measured by the same radiologist, but without reference to the original scans, and some considerable time later (approximately 1 year). The pairs of volumes for these four patients are given in table 7.2. Considerable variability was found in these estimates, with the mean error being 17%. It appeared that in some cases adjacent normal structures were included in the measurement on one occasion but not on the other. When exactly the same structures were included in the measurements, the results were consistent, and the measurements were in close agreement (see for example the measurements for patient 3 in table 7.2).

Model estimates

The estimates of sensitive tumour kill, resistance and tumour volume doubling time for each of the 8 patients are shown in table 7.3. A detailed worked example showing how the estimates were derived for patient 9 is shown in table 7.4. Initially, a guess is made for the values of the parameters (see table 7.4). The model's predictions, based on these guesses, are then compared with the actual results (by evaluating the log-likelihood as described in section 7.3). A new estimate of the four parameters is produced based on the differences between the predicted and actual

(i.e. measured) volumes (this involves using the semi-Newton algorithm described in section 7.3). This new estimate should be closer to the actual volume (and thus have a greater likelihood). This procedure is repeated until the predictions come no closer to the actual volumes (i.e. the likelihood no longer increases significantly). The likelihood for each of the volumes, given the final 'best' parameters, is given in table 7.4, along with a comparison of the predictions with the actual volumes.

In three patients the estimates for tumour volume doubling time were very long, implying a very slow growth rate. In such cases, with the tumour growing so slowly, very small volume changes would need to be detectable in order to estimate the doubling time over the short time intervals considered. Inaccuracies in the volumes measurements themselves, as previously calculated, are at least as great as these changes, making estimates of the doubling time unreliable in such cases. The doubling time in these patients has thus been assumed to be approximately 150 days, based on the estimates of others for the extremes in doubling time in SCLC (Brigham *et al.*, 1978; Tubiana & Malaise, 1979; Pearlman, 1983). This problem does not significantly affect the estimates for resistance and tumour-kill which are less sensitive to small volume differences.

The accuracy of the model's predictions (see section 7.3), measured by SDs of errors about the model's predictions (given as percentages of the tumour volumes), are also shown in table 7.3. The mean percentage SD of these errors in prediction, given as a percentage, was 6.5%, excluding the patients with only four volumes

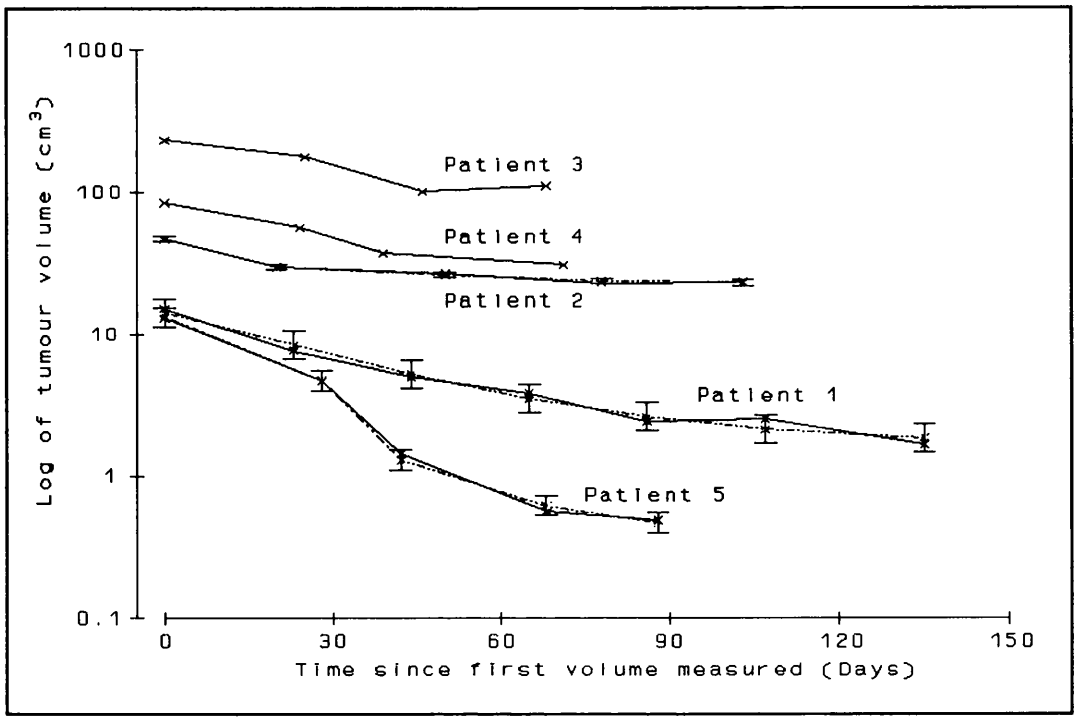


Figure 7.2 Observed tumour volumes (solid lines) and model predictions ± 1 SD (dashed lines) for patients 1-5 (observations were co-incident with predictions for patients 3 & 4).

measured, where the predictions matched the observed volumes. This percentage error is within the likely errors resulting from inaccuracies in the measurements, as described previously, and confirms that the model provides a good fit to the data. This can be seen in figures 7.2 and 7.3 which plot the observed volumes against the predictions. The chi-square goodness-of-fit tests supported this finding. There was no correlation between the percentage error and the starting tumour volume (Spearman's rank correlation coefficient = .16, P=.36). This provides some support for the assumption of a log-normal distribution of errors, since if this assumption was incorrect, some trend towards increasing or decreasing percentage errors might be expected in relation to the starting tumour volume.

The tumour volumes for the patients with more than four volumes were used to investigate the consistency of the model's predictions, and to see whether the model

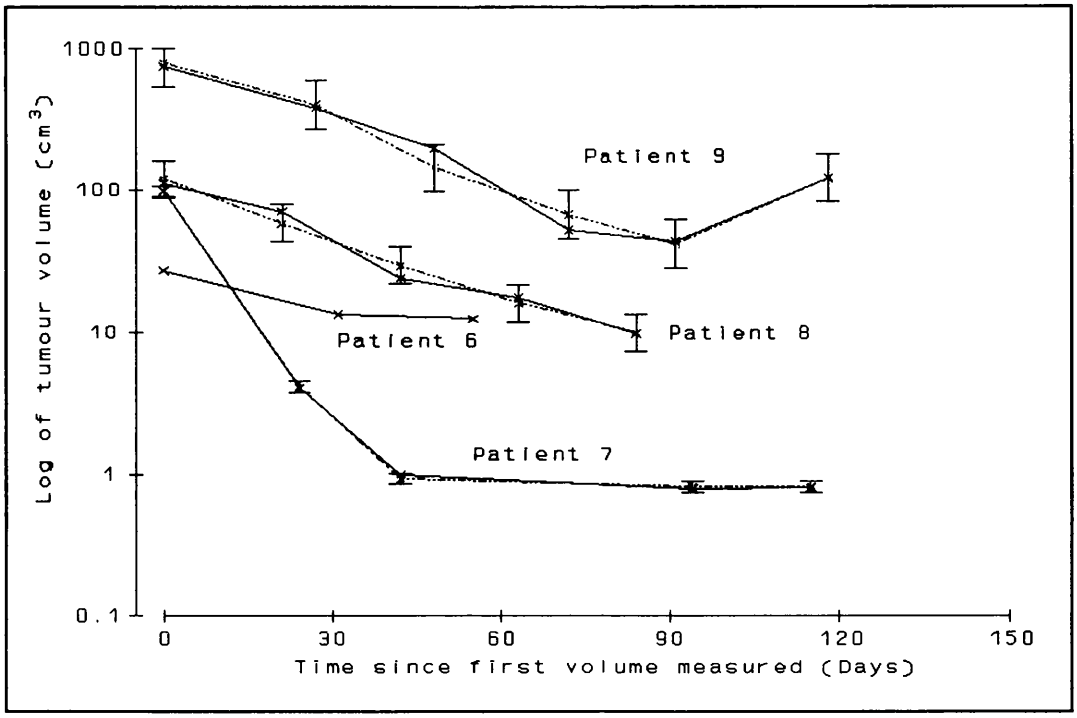


Figure 7.3 Observed tumour volumes (solid lines) and model predictions ± 1 SD (dashed lines) for patients 7-9 (patient 6 had too few observations to enable fitting of the model).

could be used in a predictive sense, for instance in deciding when to change, or to abandon a particular treatment. In these patients, the first four volumes alone were used to estimate the sensitive tumour kill, resistance and doubling time. These estimates were then used to try and predict the later volumes. These predictions, along with the actual, measured volumes are given in table 7.5. For patient 9, the predicted volumes for courses 5 and 6, using the first four volumes, bore no resemblance to the actual volumes. A further prediction of the course 6 volume, using the first 5 volumes, was also made for this patient, and this prediction is included in table 5. For the other patients the predictions are close to the actual volumes.

7.4.3 Discussion of results in Small Cell Lung Cancer

This mathematical model has two important potential uses. First, it may provide an important short-cut to obtaining information about resistance to and efficacy of chemotherapy. At present, as previously discussed, such information is only obtained from randomised trials addressing these questions, and then only by interpretation from the gross outcome measures of response duration and survival. The method described in this chapter enables these factors to be estimated for individual patients, and thus the effects of the treatments can be more easily evaluated. The patient numbers in the studies reported were insufficient to enable general conclusions to be drawn about differences in tumour kill and resistance between the different treatments. This information should, however, be obtainable from relatively small trials, although this depends on the magnitude of any differences which may occur.

The second use of this model is in predicting when to alter or stop treatment. Predictions of later volumes using earlier ones were fairly accurate, as shown by table 7.5. For patient 9, there was a clear alteration in the pattern of continued tumour reduction at the fifth volume. The reduction at this volume did not match the large reductions seen with earlier volumes. (Using the first four volumes, the fifth was predicted to be only $23cm^3$, compared with the observed value of $43cm^3$ - see table 7.5). The model detected that this lessening of the tumour-kill presaged rapid re-growth. This would have been the moment to stop treatment, or switch to a possibly non-cross-resistant alternative.

Alternative models (e.g. Birkhead *et al*, (1987)) can be considered where a proportion of the tumour is non-dividing, due, for example, to lack of vascularisation. However, this assumption was considered unnecessary, and was thought to add needless complexity in SCLC. In this tumour the monoclonal antibody Ki67, which stains cells not in the G0 phase of the cell-cycle, suggests that 60% or more of the cells are in cycle at any one time (Gatter *et al.*, 1986).

The reproducibility of the tumour volumes, especially where identical structures can be measured on each occasion, appears in this group of SCLC patients to be good, and certainly sufficient to enable estimation of the model parameters. The model appears to predict the data fairly accurately, with the average SD of errors in volume being approximately 9%.

It is interesting to note that, with the exception of patient 7, there appears to be a relationship between k, the tumour-kill, and d, the doubling time ($r=-0.89$, $p=.004$). This seems intuitively reasonable, with therapy being more effective on rapidly dividing tumours. It may be that the course 5 and 6 volumes for patient 7 represent non-dividing cells, as described.

It is likely that the doubling time of a tumour reflects a balance between the rate at which cells are proliferating and the rate of cell loss. This would not significantly affect the model's estimates or validity, since it makes assumptions only about the gross tumour volume. It may however help to reconcile the relatively slow doubling time estimated by the model with the large proportion of dividing cells found with

the monoclonal antibody Ki67, and with the relatively high cell-kills estimated and presented in table 7.3.

It is interesting that a wide variability in proportions of initially resistant tumour was seen, as suggested by Goldie and Coldman, using a model where resistance is acquired by spontaneous mutation (Goldie *et al.*, 1982).

SCLC is a highly chemo-sensitive tumour, where alterations in dose and schedule provide hope of significant, and sorely needed, improvements. A previous application of this model (Gregory *et al.*, 1988) was undertaken before this current derivation which has incorporated error distributions. Although inferences were more difficult to make, and only three volumes per patient were available, the model results helped to explain why high-dose cyclophosphamide failed to cure more patients with SCLC (Gregory *et al.*, 1988). Such explanations are needed to understand the reasons for treatment failure in SCLC, and, hopefully, may aid in the design of new and better protocols.

Table 7.1

Tumour volumes (cm^3) and times (in days since start of treatment) at which scans were performed.

		Pre course:						
Patient		1	2	3	4	5	6	7
1	Volume	15.2	7.65	4.96	3.84	2.41	2.53	1.66
	Time	0	23	44	65	86	107	135
2	Volume	46.7	29.7	27.1	23.0	23.2		
	Time	0	20	50	78	103		
3	Volume	231.7	176.8	100.9	111.0			
	Time	0	25	46	68			
4	Volume	84.2	56.3	37.3	30.9			
	Time	0	24	39	71			
5	Volume	12.9	4.66	1.43	0.56	0.49		
	Time	0	28	42	68	88		
6	Volume	27.1	13.3	12.4				
	Time	0	31	55				
7	Volume	98.4	4.1	1.0	?	0.79	0.80	
	Time	0	24	42	68	94	115	
8	Volume	111.7	70.6	24.0	17.5	9.7		
	Time	0	21	42	63	84		
9	Volume	745.0	380.5	197.2	52.6	43.4	120.8	
	Time	0	27	48	72	91	118	

? = scan not done

Table 7.2

Pairs of repeated tumour volume measurements.

	Volumes before course (cc)					
Patient	1	2	3	4	5	6
3	231.70	176.80	100.90	111.00		
	230.70	143.70	100.80	110.50		
7			1.00		0.79	0.80
			1.35		0.96	0.61
8	111.70	70.60	24.00	17.50	9.70	
	93.19	61.01	28.03	24.27	16.04	
9	745.00	380.50	197.20	52.60	43.36	120.80
	711.67	432.22	176.60	55.20	39.00	166.00

Table 7.3

Estimates of sensitive tumour-kill (k), resistance (R_0) and doubling time (d) for the 9 patients.

Patient	number of scans	Regime[a]	k(%)	R_0(%)	d (days)	likely % error[b]
1	7	OVA	46	11	>150	11
2	5	VP5	66	0.85	23	2
3	4	VP1	92	0.06	8	0
4	4	VP1	49	9	88	0
5	5	VP5	81	0.36	30	7
6	3	VP1	?	?	?	
7	5	VA	97	0.84	>150	4
8	5	OVA	59	2	92	15
9	6	VP5	90	.01	12	19

[a] OVA - Oncovin, VP16, Adriamycin
VA - VP16, Adriamycin
VP1 - VP16 given over 1 day
VP5 - VP16 given over 5 days

[b] one standard deviation of errors about model predictions (see text)

? = insufficient scans to apply model

Table 7.4

A worked example for patient 9, including a comparison of model predictions with actual volumes.

Initial guesses for the parameters for patient 9 were:

k=0.8, R_0=0.0005, d=14 days, SD (σ)=0.2, X_0=745 cm^3

(The initial log-likelihood was -48.263)

The semi-Newton maximisation routine produced the 'best' (or maximum likelihood) estimates at the following parameter values:

k=0.90, R_0=0.0001, d=11.7 days, SD (σ)=0.169, X_0=793 cc

(the maximum log-likelihood was 2.163)

The actual values and predictions were as follows

		ACTUALS		PREDICTIONS		LIKELIHOOD	
Course[a] (i)	Time (days)	X_i	Log(X_i)	u	Log(u)	L	Log(L)
0	0	745.00	6.61	793.07	6.68	2.21	0.79
1	27	380.50	5.94	400.91	5.99	2.25	0.81
2	48	197.25	5.28	143.62	4.97	0.40	-0.91
3	72	52.60	3.96	67.36	4.21	0.81	-0.21
4	91	43.60	3.78	41.54	3.73	2.27	0.82
5	118	120.80	4.79	121.21	4.80	2.36	0.86

([a] course 0 is the pre-treatment value)

Table 7.5

Model predictions of later tumour volumes from earlier tumour volumes

Patient	Number of volumes used in prediction	actuals/ predictions	Course 5	6	7
1	4	actuals	2.41	2.53	1.66
		predictions	2.50	2.11	1.88
2	4	actuals	23.2		
		predictions	17.6		
5	4	actuals	0.49		
		predictions	0.26		
7	4	actuals		0.80	
		predictions		0.83	
8	4	actuals	9.7		
		predictions	11.3		
9	4	actuals	43.4	120.8	
		predictions	23.0	11.0	
9	5	actuals		120.8	
		predictions		174.3	

Chapter 8

CONCLUSIONS

Chapter 8

CONCLUSIONS

8.1 Summary

Two mathematical models have been described and validated in this thesis, and examples of their use have been given.

The first model, the remission duration model, assumes that the progression of disease in cancer patients after a period of remission is related to the volume of residual tumour, usually resistant to therapy, remaining after treatment. It is assumed that patients are prescribed treatments which reduce the tumour volume to levels not discernable by current technology. Subsequent relapse from this state of remission is assumed to be related to the unknown resistant volume, and to the tumour growth rate. It is assumed that the distribution of resistant volumes for a population of patients given the same treatment is log-normal, and that re-growth rates are derived from a log-normal distribution of doubling times. The mathematics for estimating the parameters of these two distributions for a particular cohort of patients is presented, and applications are given.

The second model also relates to the re-growth of resistant tumour, and assumes that, during therapy, successive changes in the tumour volume can be used to estimate the

proportion of the tumour which is resistant to therapy, and the re-growth rate of the tumour. The model assumes that each course of treatment kills a constant fraction of the tumour which is sensitive to that therapy. An exponential tumour growth rate is also assumed. An application is given from a clinical trial, and it is shown how the method has the potential for choosing an appropriate time to stop the current treatment or switch to a new therapy.

These models are one way to represent the patients' progress during and following treatment and have been demonstratably successful in their validation and application to clinical trials and clinical data. No doubt other models, or modifications of those presented here, may be suggested as experience increases. However, in the case studies given in chapters 5 to 7 the inferences obtained have gained clinical credibility and have been used to help in the design of new trials.

8.2 Conclusions

It has been demonstrated that the mathematical models described in this thesis can generate new hypotheses to explain the results of clinical data, and suggest new approaches to treatment. This is shown clearly in chapter 5 with particular reference to Hodgkin's disease and locally advanced breast cancer. Clinicians have shown great interest in these models and many other applications are in progress.

The models provide the potential for matching the treatment to suit the patient, both in terms of the patient's presenting factors (as described in chapter 6), and on an individual basis in terms of their response to therapy as it progresses (as described in chapter 7). Analysis with clinicians of the feasibility of implementing this latter individualistic approach in breast cancer is in progress with Dr. R. Leonard in Edinburgh.

8.3 Discussion

Clinicians have been struggling for some 45 years with the problems of how best to use the (chemo)therapeutic agents at their disposal. Many drugs are dramatically effective at killing tumour cells, and inducing complete disease remissions. However, relapse is still the norm in many cancers. This is a very frustrating picture, since it is often felt by the clinicians designing and implementing the treatments, that relatively minor changes in dose and/or scheduling might produce striking improvements, yet such improvements rarely occur. However, evenly recently, alterations in dose and scheduling of drugs which have been in use for many years have led to great improvements in efficacy (Slevin *et al.*, 1989). One possible key to these improvements appears to lie in the exploitation of the growth kinetics of tumours, by timing the administration to hit cycling cells, and matching dose and schedule to different cancers, and ultimately to the individual patient. The principles outlined for the use of cancer therapy by Skipper and colleagues, and since built upon by many others, would enable more appropriate choice of treatment

if more information was available about tumour growth rates, acquisition of resistance, and cell-kill.

The advent of multi-drug chemotherapy, further exacerbated this situation. The choice of drug combinations and schedules is often bewilderingly large; consider for example some of the trials employed for Hodgkin's and non-Hodgkin's lymphomas (Fisher *et al.*, 1983; Skarin *et al.*, 1983; Bonadonna *et al.*, 1986; Connors & Klimo, 1988). It has taken some 15 years or more to show that the more intensive, aggressive regimes are not superior to the standard CHOP regimen for non-Hodgkin's lymphomas (Fisher *et al.*, 1992). With clinical trials only addressing the simple question of whether one regime is better than another (and this often ineffectively), progress is slow, and the literature confusing, with many puzzling and apparently contradictory results (see for example Slevin & Staquet, 1986).

A step forward in this situation can be gained from recently developed mathematical models, such as those described in this thesis. These models attempt to provide the kind of information from a trial which can lead to hypotheses to help in understanding *why* differences occurred. This information, used constructively, will contribute to better trials, and lead to new directions for research. Thus, trials should be able to progress in a more structured and rational manner.

References

A'Hern, R.P., Ebbs, S.R. & Baum, M.B. (1988). Does chemotherapy improve survival in advanced breast cancer? A statistical overview. *Br J Cancer*, **57,** 615-8.

Akanuma, A. (1983). Parameter analysis of gompertzian function growth model in clinical tumours. *European Journal of Cancer*, **14,** 681-688.

Altman, D.G. (1984). A fair trial? *Br Med J*, **289,** 336-337.

Altman, D.G. (1985). Comparability of randomised groups. *The Statistician*, **34,** 125-136.

Altman, D.G. (1991). Practical Statistics for Medical Research. Chapman and Hall: London.

Anderson, P.K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics*, **38,** 67-77.

Armitage, P. (1975). *Sequential Medical Trials*, Blackwell: Oxford.

Armitage, P. (1984). Controversies and achievements in clinical trials. *Controlled Clin Trials*, **5,** 67-72.

Armitage, P. & Berry, G. (1987). *Statistical Methods in Medical Research*, 2nd edn., Blackwell: Oxford.

Armitage, P. & Parmar, M. (1986). Some approaches to the problem of multiplicity in clinical trials. Proceedings of 13th International Biometric Conference, Seattle, Biometric Society, 1986.

Badwe, R.A., Gregory, W.M., Chaudary, M.A., Richards, M.A., Bentley, A.E., Rubens, R.D. & Fentiman, I.S. (1991). Timing of surgery during menstrual cycle and survival of premenopausal women with operable breast cancer. *The Lancet*, **337,** 1261-1264.

Barnard, N.J., Hall, P.A., Lemoine, N.J. & Kodar, N. (1987). Proliferative index in breast carcinoma determined in situ by Ki67 immunostaining and it's relationship to clinical and pathological variables. *Journal of Pathology*, **152,** 287-295.

Beale, E.M.L. (1988). Multi-dimensional Optimization. In *Introduction to Optimization*, (ed) pp. 25-36. John Wiley & Sons: Chichester.

Bell, R., Rohatiner, A.Z.S., Slevin, M.L., Ford, J.M., Dhaliwal, H.S., Henry, G., Birkhead, B.G., Amess, J.A.L., Malpas, J.S. & Lister, T.A. (1982). Short-term treatment for acute myelogenous leukaemia. *Br Med J*, **284,** 1221-1229.

Berenbaum, M.C. (1990). Direct search methods in the optimisation of cancer chemotherapy regimens. *Br J Cancer*, **61**, 101-109.

Bergsagel, D.E. (1989). Is aggressive chemotherapy more effective in the treatment of plasma cell myeloma? *Eur J Cancer Clin Oncol*, **25**, 159-161.

Birkhead, B.G. & Gregory, W.M. (1984). A mathematical model of the effects of drug resistance in cancer chemotherapy. *Mathematical Biosciences*, **72**, 59-70.

Birkhead, B.G., Gregory, W.M., Slevin, M.L. & Harvey, V.J. (1986). Evaluating and designing cancer chemotherapy treatments using mathematical models. *Eur. J. Cancer*, **22**, 3-8.

Birkhead, B.G. & Jackson, R.R.P. (1986). Randomized Trials: a perspective. In *Randomised trials in cancer: a critical review by sites*, Slevin, M. & Staquet, M. (ed) pp. 665-675. Raven: New York.

Birkhead, B.G., Jackson, R.R.P. & Salt, D. (1987a). Reliability and consistency of bone marrow assessments in acute myelogenous leukaemia. *Hematol. Onc.*, **5**, 139-145.

Birkhead, B.G., Rankin, E.M., Gallivan, S., Dones, L. & Rubens, R.D. (1987b). A mathematical model of the development of drug resistance to cancer chemotherapy. *Eur J Cancer Clin Oncol*, **23**, 1421-1427.

Black, S.D. (1979). The paradox of medical care. *J R Coll Physicians Lond*, **13**, 57.

Bloom, H., Richardson, M. & Harries, B. (1962). Natural history of untreated breast cancer (1804-1933): Comparison of treated and untreated cases according to histological grade of malignancy. *Br Med J*, **2**, 213-221.

Bloom, H.J.G. & Richardson, W.W. (1957). Histological grading and prognosis in breast cancer. *Br J Cancer*, **5**, 173-183.

BMJ Editor (1991). Obituary of Sir Austin Bradford Hill. *Br Med J*, **302**, 1017.

Bonadonna, G. & Valagussa, P. (1987). Current status of adjuvant chemotherapy for breast cancer. *Semin Oncol*, **14**, 8-22.

Bonadonna, G., Valagussa, P., Rossi, A., Tancini, G., Brambilla, C., Zambetti, M. & Veronesi, U. (1985). Ten-year experience with CMF-based adjuvant chemotherapy in resectable breast cancer. *Breast Cancer Res Treat*, **5**, 95-115.

Bonadonna, G., Valagussa, P. & Santoro, A. (1986). Alternating non-cross resistant combination chemotherapy or MOPP in stage IV Hodgkin's disease. *Annals of Int Med*, **104**, 739-746.

Brigham, B.A., Bunn, P.A. & Minna, J.D. (1978). Growth rates of small cell bronchogenic carcinomas. *Cancer*, **42,** 2880-2886.

Brown, B.W., Atkinson, E.N., Thompson, J.R. & Montague, E.D. (1987). Lack of concordance of growth rates of primary and recurrent breast cancer. *J Nat Cancer Inst*, **78,** 425-435.

Byar, D.P., Simon, R.M., Friedewald, W.T., Schlesselmann, J.J., DeMets, D.L., Ellenberg, J.H., Gail, M.H. & Ware, J.H. (1976). Randomised clinical trials: Perspectives on some recent ideas. *N Engl J Med*, **295,** 74-80.

Canner, P.L. (1987). An overview of six clinical trials of asprin in coronary heart disease. *Stats in Med*, **6,** 255-263.

Chalmers, T.C., Levin, H., Sacks, H.S., Reitman, D., Berrier, J. & Nagalingam, R. (1987). Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stats in Med*, **6,** 315-325.

Connors, J.M. & Klimo, P. (1988). MACOP-B chemotherapy for malignant lymphomas and related conditions: 1987 update and additional observations. *Seminars in Haemat*, **25,** 41-46.

Cox, D.R. (1972). Regression models and life tables. *J Royal Stat Soc (B)*, **34,** 187-220.

D'Angio, G.J., Breslow, N., Beckwith, J.B., Evans, A., Baum, H., deLorimier, A., Fernbach, D., Hrabovsky, E., Jones, B., Kelalis, P., Othersen, H.B., Tefft, M. & Thomas, P.R.M. (1989). Treatment of Wilms' tumor. Results of the Third National Wilms' Tumor Study. *Cancer*, **64,** 349-360.

D'Angio, G.J., Evans, A., Breslow, N., Beckwith, B., Bishop, H., Farewell, V., Goodwin, W., Leape, L., Palmer, N., Sinks, L., Sutow, W., Tefft, M. & Wolff, J. (1981). The treatment of Wilms' tumour. Results of the Second National Wilm's Tumour Study. *Cancer*, **47,** 2302-2311.

D'Angio, G.J., Evans, A.E., Breslow, N., Beckwith, B., Bishop, H., Feigl, P., Goodwin, W., Leape, L.L., Sinks, L.F., Sutow, W., Tefft, M. & Wolff, J. (1976). The treatment of Wilms' tumour: results of the National Wilm's Tumour Study. *Cancer*, **38,** 633-646.

Davis, P.J. & Polonsky, I. (1965). Numerical interpretation, differentiation, and integration. In *Handbook of Mathematical Functions*, Abramowitz, M. & Stegun, A. (ed) pp. 932. Dover Publications Inc.: New York.

De Vita, V.T., Simon, R.M., Hubbard, S.M., Young, R.C., Berard, C.W., Mocley, J.H., Frei, E., Carbone, P.P. & Canellos, G.P. (1980). Curability of advanced Hodgkin's disease with chemotherapy. Long term follow-up of

MOPP-treated patients at the National Cancer Institute. *Ann Intern Med*, **92,** 587-594.

Demicheli, R. (1980). Growth of testicular neoplasm lung metastases: tumour specific relation between two gompertzian parameters. *Europ. J. Cancer.*, **16,** 1603-1608.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7,** 177-188.

Early Breast Cancer Trialists' Collaborative Group (1992). Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. *The Lancet*, **339,** 1-15, 71-85.

Ellison, R.R. & Murphy, M.L. (1964). "Apparent doubling time" of leukemic cells in marrow. *Clin. Res. (abstract)*, **12,** 284.

Engelman, L. (1985). Stepwise logistic regression. In *BMDP Statistical Software Manual*, Dixon, W. (ed) pp. 330-344. Univ of California Press: London.

Findlay, P.A., Lippman, M.E., Danforth, D.J., McDonald, H., d'Angelo, T., Gorrell, C.R., Gerber, N.L., Schain, W. & Lichter, A.S. (1985). Mastectomy versus radiotherapy as treatment for stage I-II breast cancer: a prospective randomised trial at the National Cancer Institute. *World J Surg*, **9,** 671-675.

Fisher, B., Bauer, M., Margolese, R., Poisson, R., Pilch, Y., Redmond, C., Fisher, E., Wolmark, N., Deutsch, M., Montague, E., Saffer, E., Wickerham, L., Lerner, H., Glass, A., Shibata, H., Deckers, P., Ketcham, A., Oishi, R. & Russell, I. (1985). Five-year results of a randomized clinical trial comparing total mastectomy and segmental mastectomy with or without radiation in the treatment of breast cancer. *N Engl J Med*, **312,** 665-673.

Fisher, R.A. (1935). *The design of experiments*, Oliver & Boyd: Edinburgh.

Fisher, R.I., DeVita, V.T.J. & Hubbard, S.M. (1983). Diffuse aggressive lymphomas: increased survival after alternating flexible sequences of ProMACE and MOPP chemotherapy. *Annals of Internal Medicine*, **98,** 304-309.

Fisher, R.I., Gaynor, E., Dahlberg, S., Mize, E.M. & Oken, M.M. (1992). A phase III comparison of CHOP vs. m-BACOD vs. ProMACE-CytaBOM vs. MACOP-B in patients with intermediate or high-grade non-Hodgkin's lymphoma: preliminary results of SWOG-8516 (INTERGROUP 0067), the national high priority lymphoma study. *ASCO Abstract* number 1067, p315.

Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the logrank test. *Stats in Med*, **1,** 121-129.

Frei III, E. & Freireich, E.J. (1965). Progress and perspectives in the chemotherapy of acute leukemia. *Adv. Chemother.*, **2,** 269-298.

Freireich, E.J., Gehan, E., Frei, I.E., Schroeder, L.R., Wolman, I.J., Anbari, R., Burgert, E.D., Mills, S.D., Pinkel, D., Selawry, O.S., Moon, J.H., Gendel, B.R., Spurr, C.L., Storrs, R., Haurani, F., Hoogstraten, B. & Lee, S. (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukaemia: a model for evaluation of other potentially useful therapy. *Blood*, **21,** 699-716.

Furberg, C.D. & Morgan, T.M. (1987). Lessons from overviews of cardiovascular trials. *Stats in Med*, **6,** 295-303.

Gatter, K.C., Dunnill, M.S., Gerdes, J., Stein, H. & Mason, D.Y. (1986). New approach to assessing lung tumours in man. *Journal of Clinical Pathology*, **39,** 590-593.

Gerdes, J., Dallenbach, F., Lennert, K., Lemke, H. & Stein, H. (1984). Growth fractions in malignant non-Hodgkin's lymphoma as determined in situ by the monoclonal antibody Ki67. *Haematological Oncology*, **2,** 365-371.

Gerdes, J., Lelle, R.J., Pickartz, H., Heidenreich, W., Schwarting, R., Kurtsiefer, L., Stauch, G. & Stein, H. (1986). Growth fractions in breast cancers determined in situ with monoclonal antibody Ki67. *Journal of Clinical Pathology*, **39,** 977-980.

Goldie, J.H. & Coldman, A.J. (1979). A mathematical model for relating the drug sensitivity of tumours to their spontaneous mutation rate. *Cancer Treatment Reports*, **63,** 1727-1731.

Goldie, J.H., Coldman, A.J. & Gudauskas, G.A. (1982). Rationale for the use of alternating non-cross-resistant chemotherapy. *Cancer Treatment Reports*, **66,** 439-449.

Green, S.J., Fleming, T.R. & Emerson, S. (1987). Effects on overviews of early stopping rules for clinical trials. *Stats in Med*, **6,** 361-367.

Greer, S. & Watson, M. (1987). Mental adjustment to cancer: its measurement and prognostic importance. *Cancer Surv 1987*, **6,** 439 453.

Gregory, W.M., Birkhead, B.G. & Souhami, R.L. (1988). A mathematical model of drug resistance applied to treatment for small cell lung cancer. *Journal of Clinical Oncology*, **6,** 457-461.

Gregory, W.M., Reznek, R.H., Hallett, M. & Slevin, M.L. (1990). Using mathematical models to estimate drug resistance and treatment efficacy via CT scan measurements of tumour volume. *Br J Cancer*, **62,** 671-675.

Gregory, W.M., Richards, M.A. & Malpas, J.S. (1992). Combination chemotherapy versus melphalan and prednisolone in the treatment of multiple myeloma an overview of published trials. *J Clin Oncol.*, **10,** 334-342.

Grimard, L., Eapen, L., McDermot, R., Genest, P. & Nair, R. (1988). Does parasternal irradiation reduce thoracic vertebral metastases in breast cancer? *Int J Radiation Oncology Biol Phys*, **14,** 565-569.

Hall, P.A., Richards, M.A., Gregory, W.M., d'Ardenne, A.J., Lister, T.A. & Stansfeld, A.G. (1987). The prognostic value of Ki67 immunostaining in Non-Hodgkin's lymphoma. *Journal of Pathology*, **154,** 223-236.

Hanby, A.M., Hall, P.A., Rooney, N., Dennis, P., James, P., Richman, P., Buk, S., Levison, D.A. & Gregory, W.M. (1992). An inter- and intra-observer variability study on the diagnosis of lymph node biopsies. *Eur J Cancer*, **28A,** 1858-1862.

Hayward, J.L., Carbone, P.P., Heuson, J.C., Kumaoka, S., Segaloff, A. & Rubens, R.D. (1978). Assessment of response to therapy in advanced breast cancer. *Br J Cancer*, **38,** 201.

Heuser, L., Spratt, J. & Polk, H. (1979). Growth rates of primary breast cancer. *Cancer*, **43,** 1888-1894.

Hill, Sir A.B. (1952). The philosophy of the clinical trial. *N Engl J Med*, **247,** 113-115.

Hill, Sir A.B. (1962). *Statistical Methods in Clinical and Preventative Medicine*, Livingstone: Edinburgh.

Holland, J.F. (1968). Clinical studies of unmaintained remissions in acute lymphocytic leukemia. In *The M.D. Anderson Hospital, Houston, Texas: The Proliferation and Spread of Neoplastic Cells*, (ed) pp. 453-462. Williams and Wilkins Co.: Baltimore.

Hrushesky, W.J.M., Bluming, A.Z., Gruber, S.A. & Sothern, R.B. (1989). Menstrual influence on surgical cure of breast cancer. *Lancet*, **2,** 948-952.

Imanaga, H. & Nakazato, H. (1977). Results of surgery for gastric cancer and effect of adjuvant mitomycin C on cancer recurrence. *World Journal of Surgery*, **1,** 213-221.

ISIS-2 (Second International Study of Infarct Survival) (1988). Collaborative Group (1988) Randomised trial of intravenous streptokinase, oral asprin, both, or neither among 17187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet*, **2,** 349-360.

Jackson, R.R.P. (1985). The evaluation of clinical trials. *Postgraduate Med J*, **61,** 133-139.

Kalbfleisch, J.D. & Prentice, R.L. (1980). The proportional hazards model. In *The statistical analysis of failure time data*, Kalbfleisch, J.D. & Prentice, R.L. (ed) pp. 70-118. Wiley & Sons: New York.

Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Am Stat Assoc J*, **53,** 457-481.

Kincaid, D. & Cheney, W. (1991). Numerical Analysis. (ed) pp. 430-442.

Laird, A.K. (1964). Dynamics of tumour growth. *Br J Cancer*, **18,** 490-502.

Lee, B.J. (1984). Letter to the editor. *J Clin Oncol*, **2,** 245-246.

Lee, B.J., Lake-Lewin, D. & Myers, J.E. (1982). Intensive treatment of multiple myeloma. In *Controversies in Oncology*, Wiernik, P. (ed) pp. 61-79. John Wiley and Sons: New York.

Lee, T.E. (1980). In *Statistical Methods for Survival Data Analysis*, (ed) pp. 217-223. Lifetime Learning Publications: Belmont, California.

Mackenzie, A.R. (1966). Chemotherapy of metastatic testis cancer: results in 54 patients. *Cancer*, **19,** 1369-1376.

Mackillop, W.J., Ciampi, A., Till, J.E. & Buick, R.N. (1986). A stem cell model of human tumour growth: implications for tumour cell clonogenic assays. *J Nat Cancer Inst,*, **70,** 9-16.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in it's consideration. *Can Chem Rep*, **50,** 163-170.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J Nat Cancer Inst*, **22,** 719-748.

McElwain, T.J. & Peckham, M.J. (1974). Combination chemotherapy in testicular tumours. *Proceedings of the Royal Soc of Med*, **67,** 297.

McGuire, W.L., Osborne, C.K., Clark, G.M. & Knight, W.A. (1982). Steroid hormone receptors and carcinoma of the breast. *Am J of Physiology*, **243,** E99-E102.

Medical Research Council's Working Party on Lymphomas (1979). Randomised trial of two-drug and four-drug maintenance chemotherapy in advanced or recurrent Hodgkin's disease. *Br Med J*, **1,** 1105-1108.

Murphy, M.F., Metcalfe, P., Thomas, H., Eve, J., Ord, J., Lister, T.A. & Waters, A.H. (1986). Use of leucocyte-poor blood components and HLA-matched platelet donors to prevent HLA alloimmunization. *Br J Haemat*, **62,** 529-534.

Nakajima, T., Fukami, A., Ohashi, T. & Kajitani, T. (1978). Long-term follow-up study of gastric cancer patients treated with surgery and adjuvant chemotherapy with mitomycin C. *International Journal of Clinical Pharmacology and Biopharmacy*, **16,** 209-216.

Newlands, E.S., Begent, R.H.J., Rustin, G.J.S., Parker, D. & Bagshawe, K.D. (1983). Further advances in the management of malignant teratomas of the testis and other sites. *Lancet*, **1,** 948.

Norton, L. (1988). A gompertzian model of human breast cancer growth. *Cancer Research*, **48,** 7067-7071.

Norton, L. & Simon, R. (1977). Tumour size, sensitivity to therapy, and design of treatment schedules. *Cancer Treatment Reports*, **61,** 1307-1317.

Norton, L. & Simon, R. (1986). The Norton-Simon Hypothesis revisited. *Cancer Treatment Reports*, **70,** 163-169.

O'Reilly, S.M. & Richards, M.A. (1992). Clinical aspects of assessing cell proliferation. In *Assessment of Cell Proliferation in Clinical Practice*, Hall, P., Levison, D. & Wright, N. (ed) pp. 177-191. Springer-Verlag: London.

Okumura, Y., Ueda, T., Mori, T. & Kitabatake, T. (1977). Kinetic analysis of tumour regression during the course of radiotherapy. *Strahlentherapie*, **153,** 35-39.

Oliver, R.T.D. (1986). Germ cell tumours. In *Randomised trials in cancer: a critical review by sites*, Slevin, M.L. & Staquet, M.J. (ed) pp. 631-643. Raven: New York.

Pearlman, A.W. (1983). Doubling time and survival time. In *Cancer Treatment: End-point Evaluation*, Stoll, B. (ed) pp. 279-301. John Wiley and Sons: New York.

Peckham, M.J., Barrett, A. & Lieu, K.H. (1983). The treatment of metastatic germ cell testicular tumours with bleomycin, etoposide and cis-platinum (BEP). *Br J cancer*, **47,** 613-619.

Peto, R. (1987). Why do we need systematic overviews of randomised trials. *Stats in Med*, **6,** 233-240.

Peto, R. & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society, series A*, **135,** 185-198.

Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J. & Smith, P.G. (1977a). Design and analysis of randomised clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br J Cancer*, **35,** 1-39.

Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J. & Smith, P.G. (1977b). Design and analysis of randomised clinical trials requiring prolonged observation of each patient: I introduction and design. *Br J Cancer*, **34,** 585-612.

Pocock, S.J. (1984). *Clinical trials. A practical approach.*, Wiley: New York.

Pocock, S.J. & Hughes, M.D. (1990). Estimation issues in clinical trials and overviews. *Stats in Med*, **9,** 657-671.

Price, P., Hogan, S.J., Bliss, J.M. & Horwich, A. (1990a). The growth rate of metastatic nonseminomatous germ cell testicular tumours measured by marker production doubling time--II. Prognostic significance in patients treated by chemotherapy. *Eur J Cancer*, **26,** 453-457.

Price, P., Hogan, S.J. & Horwich, A. (1990b). The growth rate of metastatic non-seminomatous germ cell testicular tumours measured by marker production doubling time--I. Theoretical basis and practical application. *Eur J Cancer*, **26,** 450-453.

Rees, J.K.H., Gray, R.G., Swirsky, D. & Hayhoe, F.G.J. (1986). Principal results of the medical research council's 8th acute myeloid leukaemia trial. *Lancet*, **8518,** 1236-1241.

Richards, M.A., O'Reilly, S.M., Howell, A., George, W.D., Fentiman, I.S., Chaudary, M.A., Crowther, D. & Rubens, R.D. (1990). Adjuvant cyclophosphamide, methotrexate, and fluorouracil in patients with axillary node-positive breast cancer: an update of the Guy's/Manchester trial. *J Clin Oncol*, **8,** 2032-2039.

Rubens, R.D., Bartelink, H., Engelsman, E., Hayward, J.L., Rotmensz, N., Sylvester, R., Van, d.S., E., Papadiamantis, J., Vassilaros, S.D., Wildiers, J. & Winter, P.J. (1989). Locally advanced breast cancer: the contribution of cytotoxic and endocrine treatment to radiotherapy. *Eur J Cancer Clin Oncol*, **25,** 667-678.

Rubinow, S.I. & Lebowitz, J.L. (1976). A mathematical model of the acute myeloblastic leukemic state in man. *Biophysical J*, **16,** 897-910.

Sarrazin, D., Le, M.G., Arriagada, R., Contesso, G., Fontaine, F., Spielmann, M., Rochard, F., Le-Chevalier, T. & Lacour, J. (1989). Ten year results of a randomised trial comparing a conservative treatment to mastectomy in early breast cancer. *Radiotherapy Oncology*, **14,** 177.

Senie, R.T., Rosen, P.P., Rhodes, P. & Lesser, M.L. (1990). Prognosis of primary breast cancer patients in relation to time of diagnostic surgery during the menstrual cycle. *Breast Cancer Res Treat (abstract)*, **16,2,** 146.

Shackney, S.E. (1970). A computer model for tumour growth and chemotherapy, and its application to L1210 leukemia treated with cytosine arabinoside (NSC-63878). *Cancer Chemotherapy Reports*, **54,** 399-429.

Shackney, S.E., McCormack, G.W. & Cuchural Jnr, G.J. (1978). Growth rate patterns of solid tumours and their relation to responsiveness to therapy. *Annals of Internal Medicine*, **89,** 107-121.

Silvey, S.D. (1970). The likelihood-ratio test and alternative 'large-sample' equivalents of it. In *Statistical Inference*, Silvey, S. (ed) pp. 108-122. Penguin: Baltimore.

Simes, R.J. (1986). Publication Bias: the case for an international registry of clinical trials. *J Clin Oncol*, **4,** 1529-1541 (1986).

Simon, R.J. (1986). Confidence intervals for reporting results of clinical trials. *Annals of Int Med*, **105,** 429-435.

Skarin, A.T., Canellos, G.P. & Rosenthal, D.S. (1983). Improved prognosis of diffuse histiocytic and undifferentiated lymphoma by use of high dose methotrexate alternating with standard agents (M-BACOD). *J Clin Oncol*, **1,** 91-98.

Skipper, H.E. & Perry, S. (1970). Kinetics of normal and leukemic leukocyte populations and relevance to chemotherapy. *Cancer Research*, **30,** 1883-1897.

Skipper, H.E., Schabel, F.M.J. & Wilcox, W.S. (1964). Experimental evaluation of potential anticancer agents. XIII. On the criteria and kinetics associated with 'curability' of experimental leukemia. *Cancer Chemotherapy Reports*, **35,** 1-111.

Skipper, H.E., Schabel, F.M.J. & Wilcox, W.S. (1967). Experimental evaluation of anticancer agents. XXI. Scheduling of arabinosylcytosine to take advantage of its S-phase specificity against leukemia cells. *Cancer Chemotherapy Reports*, **51,** 125-165.

Slevin, M.L. (1986). Ovarian Cancer. In *Randomised trials in cancer: a critical review by sites*, Slevin, M.L. & Staquet, M.J. (ed) pp. 385-416. Raven: New York.

Slevin, M.L., Clark, P.I., Joel, S.P., Malik, S., Osborne, R.J., Gregory, W.M., Lowe, D.G., Reznek, R.H. & Wrigley, P.F.M. (1989). A randomised trial to evaluate the effect of schedule on the activity of etoposide in small cell lung cancer. *J Clin Oncol*, **7,** 1333-1340.

Slevin, M.L. & Staquet, M.J. (1986). *Randomised trials in cancer: a critical review by sites*, Raven: New York.

Somers, R., Henry-Amar, M., Meerwaldt, J.K. & Carde, P. (1990). *Treatment strategy in Hodgkin's disease*, John Libbey & Company: London.

Steel, G.G., Adams, K. & Stanley, J. (1976). Size-dependence of the response of Lewis lung tumours to BCNU. *Cancer Treatment Reports*, **60,** 1743-1748.

Sullivan, P.W. & Salmon, S.E. (1972). Kinetics of tumour growth and regression in IgG multiple myeloma. *Journal of Clinical Investigations*, **51,** 1967-1973.

Tsukuma, H., Mishima, T. & Oshima, A. (1983). Prospective study of "early" gastric cancer. *Int J Cancer*, **31,** 421-426.

Tubiana, M. & Malaise, E. (1979). Combination of radiotherapy and chemotherapy: Implications derived from cell kinetics. In *Lung Cancer: Progress in Therapeutic Research*, Muggia, F. & Rosenezwig, M. (ed) pp. 51-61. Raven: New York.

Tubiana, M., Pejovic, M.H., Koscielny, S., Chavaudra, N. & Malaise, E. (1989). Growth rate, kinetics of tumour cell proliferation and long-term outcome in human breast cancer. *Int J Cancer*, **44,** 17-22.

Tucker, J., Murphy, M.F., Gregory, W.M., Lim, J., Rohatiner, A.Z.S., Waters, A.H. & Lister, T.A. (1989). Apparent removal of graft-versus-leukaemia effect by the use of leucocyte-poor blood components in patients with acute myeloblastic leukaemia. *Br J Haemat (letter)*, **73,** 572-574.

VanderSchueren, E. & VanDongen, J.A. (1988). Management of early stage breast cancer - current status of treatment: workshop report. *Eur J Cancer Clin Oncol*, **4,** 89-93.

Vaughan, W.P., Karp, J.E. & Burke, P.J. (1984). Two-cycle timed sequential chemotherapy for adult acute nonlymphocytic leukemia. *Blood*, **64,** 975-980.

Veronesi, U., Banfi, A., Salvadori, B., Luini, A., Saccozzi, R., Zucali, R., Marubini, E., Del Vecchio, M., Boracchi, P., Marchini, S., Merson, M., Sacchini, V., Riboldi, G. & Santoro, G. (1990). Breast conservation is the treatment of choice in small breast cancer: long-term results of a randomized trial. *Eur J Cancer*, **26,** 668-70.

Vugrin, D., Whitmore, W.E.J. & Golbey, R.B. (1983). VAB-6 combination of chemotherapy without maintenance in treatment of disseminated cancer of the testis. *Cancer*, **51,** 211.

Wagstaff, J., Gregory, W.M., Swindell, R., Crowther, D. & Lister, T.A. (1988). Prognostic factors for survival in stage IIIB and IV Hodgkin's disease: A

multivariate analysis comparing two specialist treatment centres. *Br J Cancer*, **58,** 487-492.

Young, R.C., Canellos, G.P., Chabner, B.A., Schein, P.S. & DeVita, V.T. (1973). Maintenance chemotherapy for advanced Hodgkin's disease in remission. *Lancet*, **1,** 1339-1343.

Yusuf, S., Collins, R. & Peto, R. (1984). Why do we need some large, simple randomised trials? *Stats in Med*, **3,** 409-420.

Yusuf, S., Peto, R., Lewis, J., Collins, R. & Sleight, T. (1985). Beta blockade during and after myocardial infarction: An overview of the randomised trials. *Progress in Cardiovascular Diseases*, **27,** 335-371.