STATISTICS IN MEDICINE, VOL. 10, 1789–1817 (1991)

A DECADE OF PROGRESS IN STATISTICAL METHODOLOGY FOR CLINICAL TRIALS

RICHARD SIMON

National Cancer Institute Building EPN, Room 739, Bethesda, MD 20892, U.S.A.

SUMMARY

Clinical trials played a dominant and expanding role in the evaluation of new treatments during the decade of the 1980s. There were major improvements in the quality of clinical trials in many medical fields. There were also important developments in the methodology of designing, monitoring, conducting, analysing, reporting and interpreting clinical trials. This paper attempts to review some of these developments. A comprehensive review is beyond the abilities of any one individual. Consequently, this paper attempts to offer a broad stroke description of this area and to highlight specific topics of importance based on my particular experience. An extensive, but non-comprehensive bibliography is included to provide entry points to the literature of methodologic developments for clinical trials in the 1980s.

1. INTRODUCTION

It is important for professions, as well as individuals, to pause and look back over where they have been. This helps us evaluate where we are and where we want to go. This session on Statistical Methodology for Medicine in the 1980's at the 1990 Joint Statistical Meeting is such an opportunity. I was honoured to be asked to contribute the clinical trials segment of this review, and I quickly agreed. It was only later that I realized the magnitude of the task. I reviewed the past ten years of journals which feature papers on clinical trial methodology and was impressed by two things: first, by the number, quality and variety of articles dealing with statistical methods for clinical trials, and second, by how much of this literature I had not read. The first feature impressed on me how many talented people were working in the area of clinical trials. Moreover, these papers do not reflect the efforts of an even greater number of talented and dedicated statisticians who are involved in doing clinical trials, but not necessarily publishing methodologic articles. The extent of this literature, and my unfamiliarity with much of it also impressed upon me the futility of my attempting a comprehensive review. It also became clear that ranking the literature with regard to importance would not be useful. The usefulness of statistical tools can only be determined by our profession as a whole over time, and not enough time has elapsed to judge many of the methods. This is just as well, because the improvement and extension of clinical trials is not an olympic sport. I believe that we should emphasize our common objectives, not our competitive tendencies.

My purpose here is to provide a broad-stroke description of some of this extensive literature and to highlight specific areas of importance based on my particular experience. A detailed review of each such area would itself require an article the length of this one. I have tried to provide entry points to major components of this body of work, but the bibliography included here is certainly not complete.

2. SEQUENTIAL DESIGN AND MONITORING

The literature on sequential medical trials began in the 1950's and was highlighted by the first edition of Peter Armitage's book in 1960.[1] The literature continued to grow for the next thirty years, but the 1980's represent a period of dramatic increase in the utilization of this methodology and extension of it to satisfy the practical demands of complex, multi-centre clinical trials. Clinical trial data almost always becomes available sequentially over time because of the staggered entry of patients. However, staggered patient entry in clinical trials does not correspond well with the problem of industrial reliability testing of parts and is one reason why much of the sequential design literature based on reliability testing has had limited applicability to clinical trials. Clinical trials, particularly multi-centre trials, are logistically complex with delays in information flow and heterogeneity in the patients studied. These factors also limited the applicability of designs based on continuous analysis, parametric models and quickly observed responses. Nevertheless, there has always been a strong motivation for the use of sequential methodology in clinical trials. For randomized clinical trials of serious diseases there is the ethical imperative to stop entering patients and to report the results as soon as the superiority of one treatment is compelling. There is also the tradition of physicians re-analysing accumulating data and reporting 'significant' findings as soon as a nominal probability level $p < 0{\cdot}05$ is seen. The erroneous conclusions filling the medical literature from such practices has created a strong need for the use of sequential methods. For diseases where the pace of progress is limited by the availability of patients for clinical trials, and not by a shortage of innovative treatments to test, sequential methods are also motivated by the opportunity cost of continuing trials beyond the point where a clinically significant benefit of the new treatment is unlikely. This consideration is also important when the experimental treatments are more toxic than the controls.

During the 1980's there was increased application of group sequential methods to clinical trials, the extension of these methods for use with survival data and covariates, the development of new boundaries for early termination and the development of the use-function approach for providing flexibility in the timing of group-sequential analyses.[2-31] The group sequential approach, developed in the mid-seventies,[18] was an important practical development because data monitoring committees generally meet at discrete time intervals, and continuous analysis is not practical for multi-centre trials. The group sequential approach specifies that interim analyses will be done at a limited number of pre-specified points during the course of the clinical trial. Generally the points are equally spaced with regard to available information about treatment effect on the main endpoint although the use-function method eased that restriction.

Group sequential methods were originally developed for two-sided test situations; early termination only occurred when one treatment was shown to be significantly worse than the other with regard to the primary endpoint. Most clinical trials compare a new treatment to a control which may be a placebo, no therapy or a standard treatment. Whereas the use of one-sided or two-sided significance levels is controversial even for such clinical trials, the medical decision-making process is often asymmetrical, in contrast to the usual group-sequential boundaries. Often, for example, the new treatment will be rejected unless it is shown significantly better than the control. Either because of the additional toxicity or inconvenience of the new treatment or because of the monetary or opportunity costs involved, it may be desirable to terminate the study once it is determined that the new treatment is not a clinically significant improvement. Group-sequential designs[32-40] and two other general approaches for dealing with treatment versus control clinical trials were developed in the 1980's. These general methods are stochastic curtailment and repeated confidence intervals.

Clinical investigators have long asked questions like 'What is the probability of rejecting (or not rejecting) the null hypothesis at the end of the study given the current interim data?'. Such

conditional acceptance or rejection calculations have been found useful by data monitoring committees and also for decision making in clinical trials that were not initially planned with sequential monitoring. Even with a group sequential design, it sometimes occurs that patient accrual decreases to the extent that the original objectives cannot be met and tradeoffs between further accrual and extended follow-up are facilitated by conditional power calculations. The statistical tools based on these calculations have been called stochastic curtailment methods.[41-47]

The central idea is that if the conditional probability of rejection of the null hypothesis is large even when calculated under the null, then early termination may be warranted. Similarly, if the conditional probability of acceptance of the null hypothesis is large even when calculated under the alternative of original interest, then early acceptance of the null may be warranted. Lan *et al.*[46] showed that even if results are continuously monitored in this way and the trial stopped when the conditional probability is at least γ, then the true type I error is less than α/γ and the type II error is less than β/γ. These bounds are quite conservative for non-continuous monitoring and hence γ values of 0·8 have little effect on planned error probabilities. The fact that the bounds hold for continuous monitoring means that the method can be used even for unplanned data-selected interim analyses.

Another general method of sequential monitoring is that of repeated confidence intervals.[48-52] Jennison and Turnbull showed how to generate repeated confidence intervals for a treatment effect from group sequential boundaries. The repeated confidence intervals are valid as simultaneous confidence intervals for the treatment effect regardless of the stopping 'rule' used for the clinical trial. The confidence intervals can themselves be used as stopping rules; for example, stop when the interval excludes zero or the smallest difference of clinical significance. But the intervals also provide useful information about treatment effect for the difficult decisions faced by data monitoring committees when no clear stopping rule was specified. In cases where a sequential monitoring boundary is specified and employed, there are alternative approaches to interval estimation of the treatment effect after stopping occurs which give narrower intervals than the repeated intervals.

Statistical methods are of no value unless they are used. Two-stage designs are simple to use, particularly for single institution clinical trials performed without data monitoring committees or ongoing statistical involvement. A reasonable proportion of the efficiency of fully sequential designs is achievable in a two-stage design. More importantly, however, even a two-stage design provides a specified structure for the monitoring and reporting of results. The lack of such structure and the naive interpretation of p values has led to much confusion in the medical literature. In the past decade a variety of useful two-stage designs were developed for randomized clinical trials.[53-60] For example, the simple design of Ellenberg and Eisenberger[55] was developed for early acceptance of the null hypothesis in treatment versus control trials with binary responses. The maximum sample size is planned based on fixed sample considerations. The single interim analysis is conducted after n patients per treatment are evaluable. The value of n is chosen so that the probability of the treatment difference favouring the control is 0·05 when the alternative hypothesis of interest is true. The trial is terminated early and the null accepted if such a treatment difference is obtained. Ellenberg and Eisenberger showed that such an approach causes very little reduction in statistical power under the alternative hypothesis of interest, and it reduces the expected sample size by about 25 per cent under the null hypothesis. Others have extended and optimized this design[58,60] or developed related designs.[53,54] The design is easily extended to also permit early rejection of the null hypothesis.

Two-stage designs have also been developed which combine the functions of selection among experimental treatments and comparison to a control. In many areas of medicine there are too many new treatments to evaluate all of them in randomized clinical trials. The usual approach in

cancer is for various institutions to conduct non-randomized 'pilot' studies of a variety of new treatments and then to select one for evaluation in a randomized clinical trial compared to a control regimen. One difficulty with this approach is that the results of the pilot trials are difficult to evaluate because they are conducted at different places; hence results may be confounded with institution effects or differences in patient prognoses. Thall *et al.*[56,57] have developed two-stage designs in which the first stage is used to screen new treatments and the second stage is used to compare the promising treatments selected in the first stage to a control. With some of these designs the control is included among the randomized treatments in the first stage. Whitehead[59] has also considered a two-stage process for selecting a treatment to be evaluated in a randomized clinical trial compared to the control. These designs begin to address the strategic aspects of clinical experimentation rather than the tactical aspects of doing a single clinical trial reliably, efficiently and ethically. Strategy is important because the goal of discovering and documenting treatments that work depends on good strategy as well as good tactics. There is currently interest in selection/testing designs for the development of AIDS treatments where short-term endpoints may be used for first-stage selection and longer-term endpoints for final evaluation.

During the 1980's there has also been an increase of interest in Bayesian methods for clinical trials. To some extent the traditional acrimonious debate has continued, a debate couched in such general terms and based on such poor communication that it was not very useful. In the past, some proponents of the Bayesian position had such limited connection with major clinical trials that their arguments rang hollow. This aspect of the debate seems to be changing, however. Highly respected and experienced clinical trial statisticians are exploring Bayesian ideas and raising the level of discussion. Bayesian methods have been further developed for the design and sequential monitoring of randomized clinical trials.[61-74] Such methods are not yet widely used in clinical trials and only the future will tell whether this will change.

One final area of sequential design and analysis that has received attention in the 1980's is that of point and interval estimation of the treatment effect following a sequentially monitored clinical trial.[75-89] The size of a treatment effect is important for medical decision making. For example, an expensive or toxic treatment that produces only a modest improvement on a primary endpoint may not be recommended. There are often multiple endpoints and the size of treatment effects on each must be weighed in clinical decision making for individual patients. If we use a sequential monitoring rule that terminates the study early for large observed treatment differences, then the maximum likelihood estimator of the treatment effect will be biased; we will tend to over-estimate the treatment effect. The size of the bias depends on the monitoring boundaries used. If the boundaries are 'conservative', that is, early termination only occurs for very extreme early differences, then the degree of bias will not be large. The standard confidence intervals for the treatment effect are not guaranteed to have the correct coverage probabilities when used with sequential monitoring. Because of the growing recognition of the importance of estimation in clinical trials, a variety of point estimates and exact confidence intervals have been developed based on the sequential monitoring boundaries employed. Whereas there is no uniquely best point or interval estimator to use with sequential monitoring, we now have a body of useful methodology for examining the degree of distortion of our conventional estimates induced by sequential monitoring.

3. SAMPLE SIZE PLANNING

An effective clinical rial must ask an imporant question and provide a reliable answer. A major determinant of the reliability of the answer is the sample size of the trial. Trials of inadequate size may cause contradictory and erroneous results and thereby lead to the inappropriate treatment

of patients. They also divert limited resources away from useful applications and cheat the patients who participated in what they thought was important clinical research. Sample size planning is, therefore, a key component of clinical trial methodology.

The past decade produced a large volume of methodology useful in planning the size and duration of randomized clinical trials.[90-139] In the cancer field the methods of Rubinstein *et al.*,[126] Bernstein and Lagakos,[90] and Schoenfeld[127] are widely used for planning trials where survival is the endpoint. Other contributions to this problem take into account factors such as non-compliance, non-uniform rate of patient entry, time dependent losses to follow-up, non-standard survival distributions etc. The method of Sposto and Sather[131] is often used to plan survival studies in which there is a cure-rate, and others have also developed useful approaches. There has been a re-evaluation of some of the sample-size formulas previously used for the two-sample binomial problem. Many of those methods were found to be inaccurate, but theoretical arguments over conditional or unconditional tests continued. Whereas the simple chi-square test without continuity correction may have the correct type I error rate α for independent binomial sampling, its size conditional on the observed total number of successes can exceed α. This disturbs some people because there are 'recognizable subsets' of the sample space for which the size is incorrect. Others attack Fisher's 'exact' test as being inappropriate when the total number of successes is not fixed by design. Conditioning on the total number of successes is, however, a device for developing a test statistic whose null distribution is independent of the nuisance parameter. Also, it is not clear that independent binomial sampling rather than hypergeometric sampling is the more appropriate framework. These arguments continue. For cancer trials, however, the methods based on the power function of Fisher's exact test, and the excellent approximations available such as that developed by Ury and Fleiss[136] are widely used. Recently 'exact' unconditional methods have been developed by Suissa and Shuster.[134]

Most clinical trials have been planned within the hypothesis testing framework of Neymann and Pearson. A null hypothesis is tested at significance level α and the sample size is determined to provide power $1 - \beta$ for rejecting the null when a specified alternative hypothesis is true. There is one principle of clinical trials that is more fundamental than the liklelihood principle, the conditionality principle or the repeated sampling principle. That is the principle that everyone wants to do their own thing. This leads investigators and statisticians to sometimes specify an alternative hypothesis that is unrealistically extreme so that the required sample size looks feasible for that institution or group of institutions. This practice is devastating for clinical research for the reasons given above. There is a great need for innovation in exploring various approaches to treatment, but the place for such diversity is generally earlier than the phase III trial stage. The usual hypothesis testing approach has also been misleading for the development and reporting of therapeutic equivalence clinical trials.[48,91,113,114,133] In a therapeutic equivalence trial, the new treatment will be accepted if it appears equivalent to the control treatment with regard to the primary endpoint. For cancer trials this is usually because the control treatment is considered effective and the new treatment has some advantage with regard to toxicity, morbidity, cost or convenience. Hence, the medical decision making structure is different for therapeutic equivalence trials compared to the more usual clinical trial. In a therapeutic equivalence trial planned and analysed in the usual way, failure to reject the null hypothesis means adoption of the new treatment. If the power is poor for clinically important treatment effects Δ, then an inferior treatment will be accepted. When death is the endpoint, only very small reductions in treatment effectiveness are usually acceptable in exchange for reductions in side-effects; hence small values of Δ and β are necessary. Often somewhat larger values of α are acceptable. This is often not recognized in the planning and analysis of therapeutic equivalence trials, however. Consequently, inadequately sized trials and misleading conclusions can result.

The planning of therapeutic equivalence trials has received increasing attention in the past decade. Several variations of the usual Neymann–Pearson approach have been developed. My own preference, however, remains the use of confidence intervals in the planning and reporting of such studies.[48,114] Confidence intervals for the treatment difference make it very explicit how definitive the trial is. In reporting the results of a significance test, the important information is hidden in often unstated statistical power. Also, statistical power takes no account of the results actually obtained.

Statisticians have been very effective in developing methodology for determining sample size as a function of the difference to be detected. Clinical trials in the 80's benefited from these developments, and even more importantly, from a growing recognition of the need for larger sample sizes. Sometimes the most important role of the statistician is in arguing persuasively that a meaningful independent trial is not feasible and that the institution or group should participate in an ongoing study or collaborate with others in a new one. Statisticians should also become more active in estimating the degree of difference that can be realistically expected for a clinical trial based on previous studies in that disease, and perhaps, when these are lacking, based on subjective opinions of investigators and physicians not involved in the proposed trial.

4. TREATMENT ASSIGNMENT METHODS

In cancer therapeutics during the 1970's there was much controversy over the use of historical controls in clinical trials. Were they 'valid'? Was randomization needed? During the past decade much of this controversy seems to have receded. It has become more widely recognized that both non-randomized and randomized clinical trials have appropriate roles, but that these roles are different.[156] For definitive evaluation of whether a new treatment is better than the standard, randomization is almost always needed. The exceptions are those cases where the prognosis of patients with the control treatment is uniformly bad and the new treatment has a dramatic long term effect. However, we cannot do randomized clinical trials of all candidate treatments, and the best ideas for new treatments often come from individuals who do not have access to enough patients to do a major randomized clinical trial and who do not have enough influence to convince others to study their treatment in the absence of some promising 'pilot' data. Non-randomized 'pilot' studies of innovative treatments provide the data supporting regimens worthy of investigation in major randomized clinical trials. There are, of course, examples where treatments are accepted into medical practice on the basis of results of 'pilot' studies without ever being evaluated in randomized clinical trials. Also, 'pilot' evaluations of new treatments are often performed poorly, with no controls at all or no serious attempt to evaluate whether the historical controls used are similar in prognosis to those given the new treatment. Nevertheless, the concept of historically controlled screening studies followed by major randomized clinical trials seems to have become firmly established and accepted in the 1980's by the medical oncology community.

A second controversy concerning treatment assignment methods was whether to and how to stratify the randomization procedure with regard to prognostic variables.[140–161] Physicians like to see the treatment groups balanced with regard to prognostic variables. They trust such balance more than they do complex covariate adjustments. In the late 1970's and early 1980's several dynamic stratification methods were developed that permitted one to effectively balance treatment groups with regard to more covariates than did conventional permuted block methods. The use of such methods, and stratification itself, was however challenged on the basis of (a) complexity that might limit patient accrual; (b) limited value in increasing statistical power for detecting treatment effects; (c) causing difficulties in the analysis of results. Stratified randomization, including the use of dynamic stratification methods, is used throughout the U.S.A. and Europe by major national and international cancer co-operative groups. There is no indication

that it hinders accrual. Studies by Begg and Kalish[142] for logistic regression analysis indicate that compared to balanced designs, complete randomization is frequently inefficient, and will occasionally result in a trial that is very inefficient. Halpern and Brown[146] studied analysis issues and concluded that 'Classical (analysis) procedures can generally be trusted unless trends in outcome of low time-frequency are quite clearly manifest in the stratifying variables.' In the latter case, the classical analysis can be unduly conservative. Others who have studied the problem have concluded that the increase in statistical power compared to analytic covariate adjustment is often small but that there are no serious analytical difficulties. As indicated above, there are advantages to stratification other than overall statistical power. Senn[155] showed that even if pure randomization is used in a very large clinical trial, covariate adjustment of the treatment difference can be very important, and that lack of statistical significance of the imbalance is not an adequate indicator of when there is a problem.

The decade of the 80's saw a continuation of research into outcome adaptive treatment assignment methods.[162-171] These are methods which attempt to assign a greater proportion of patients to the treatment which is doing better with regard to interim results. There is an immense volume of literature on such generalized two-arm bandit strategies yet the research has found almost no applications in real clinical trials. In fact, the past decade produced the first known application of this approach to a clinical trial. It also produced a major controversy over what we know and when we know it.[170] Unfortunately, much of the acrimony concerning the ethics of clinical experimentation and criticism about the lack of use of adaptive assignment methods derives from a lack of experience with major clinical trials. I am not optimistic about the usefulness of this area of research. We approach patients to enter a clinical trial because we believe that there is genuine uncertainty about which treatment is better; we may have our biases but have learned not to trust them. The decision about when to stop entering patients is difficult and is best placed in the hands of a properly constituted data monitoring committee. This permits the individual physician to deal with his patients honestly and places the burden of evaluating interim results in the hands of a group equipped to deal with it. I do not find it attractive to approach a patient saying that I do not know which treatment is better, but treatment A is doing better therefore I will give you a greater than 50 per cent chance of getting it. I question whether meaningful clinical trials based on adaptive assignment are compatible with fully informed consent.

5. PHASE II CLINICAL TRIALS

The randomized phase III clinical trial has traditionally received the most statistical attention. In the field of cancer, at least, the results of non-randomized phase II clinical trials determine what phase III trials will be performed. Hence, there are more phase II clinical trials than phase III trials and the former are very important. The phase II trials are of two basic types. Early phase II trials determine whether a single agent has anti-tumour activity against a particular type of cancer and provide a rough estimate of the level of that activity. Second there are phase II trials to determine whether the level of activity of a particular combination of active agents is sufficiently promising to warrant a randomized phase III evaluation of it against the standard treatment.

The first type of trial has received considerable statistical attention in the 1980's.[172-183] Fleming's one sample group sequential design[174] has been extensively used by cancer investigators. More recently Simon[179] and Chang *et al.*[173] introduced optimized versions of this type of group sequential design. The optimization minimized expected sample size subject to upper bounds on the error probabilities. Schaid *et al.*[177] studied the inclusion of a phase II treatment arm as part of a phase III clinical trial and Simon *et al.*[180] discussed the advantages of randomized phase II trials for selecting the most promising of several new agents. Herson and

Carter[176] described calibrated phase II trials that included randomization of some patients to a standard treatment in order to check that the patient population is capable of responding to an active treatment. Herson[175] proposed using predictive probability for early termination of phase II trials and Sylvester[181] introduced a decision theoretic approach for determining the size of phase II trials in terms of likelihood of patient benefit. Whitehead[183] introduced a Bayesian method for establishing the size of phase II trials by balancing the opportunity cost of missing a good treatment by not studying it against the cost of missing it by studying it with too small a sample size.

The methods described above are useful for designing phase II trials of new drugs. They are not entirely adequate for the second type of trial described above, however. That type of study requires that explicit consideration be given to results with the standard treatment. The comparative nature of phase II trials of combination regimens is often ignored. Papers by Makuch and Simon,[115] Dixon and Simon[95] and Emrich[99] have considered this from the point of view of sample size. A recent paper by Thall and Simon[182] provides a model for determining when historical controls are sufficient for this phase II purpose, and when concurrent controls or some combination of the two is needed.

During the 1980's there was also statistical methodology developed for the design of dose ranging clinical trials.[184-195] The number and importance of such trials for cancer treatment has increased greatly with the development of biological drugs that protect normal tissues. Dose ranging trials are varied and challenging statistically. It is likely that such studies will receive increased attention in the 1990's.

6. OTHER DESIGN ISSUES

The crossover design received considerable attention during the 1980's.[196-199] The decade began with Brown's[196] description of the position of the FDA's Biometric and Epidemiologic Methodology Advisory Committee criticism of the two-period crossover design. The criticism was essentially that the efficiency of the design was based on the assumption that carry-over effects, if they exist, are the same for each treatment. The sample size needed for adequate statistical power to test this assumption would itself eliminate the efficiency of the design. Willan and Pater[199] pointed out that even if treatment dependent carry-over effects do exist, so long as their difference is of the same sign as that of the difference in main treatment effects and of a smaller magnitude, then valid analysis with efficiency gains were still possible from the crossover design. During the 1980's many variations on the simple crossover were introduced to circumvent this problem and much of this work was reviewed by Matthews.[198] Fleiss[197] critiqued these attempts in 1989 and concluded that they had failed as general solutions because they were based on new assumptions that required verification for the particular clinical situation and treatments under study. He recommended: 'The two-treatment crossover design should be adopted only if it is known that average disease severity will not change markedly over the course of the trial, and if equality or near-equality of carryover effects may be assumed *a priori*.'

For situations where disease severity was stable and carryover effects were negligible, '*N* of one' designs were introduced.[200-202] These designs involve the re-randomization of two or more treatments to subsequent treatment periods of a single patient. Inferences about relative treatment effects are specific to that patient. Such designs are attractive because of their specificity and because they broaden the scope of the clinical trial to individual physicians or even individuals treating themselves. As for the crossover design, however, the validity of results will depend on assumptions which often cannot be adequately tested by the data.

The two by two factorial design also offers efficiency and has been recommended as a way of answering two therapeutic questions at the price (in sample size) of one.[203-206] If, however, we establish a sample size sufficient for testing main effects under the assumption of no interaction, then we will not have a large enough sample size to detect interactions of the same magnitude. If the endpoints for evaluating the two factors are distinct and there is little chance of an interaction, or if the magnitude of the interaction can be assumed small relative to the size of the main effects, or if the likelihood of both main effects being clinically significant is small, then the factorial design will be quite attractive. In many cases, however, the chance of an important interaction is substantial; for example, treatment A or treatment B may each be useful individually but one may be redundant in the presence of the other. In such cases the use of a factorial design pre-judges an imporant therapeutic question that can only be answered with a larger sample size.

The methodology of response surface designs has been extensively developed for application to combination drug trials in the 1980's.[207] This methodology has received limited use for clinical trials as yet but may find more applications in the 1990's.

A great deal of methodologic work during the 1980's was devoted to being able to utilize more complex endpoints for clinical trials.[208-222] Great advances were made in the design and analysis of studies with repeated measurements on the same individuals.[222] Some of the new methodology is much more flexible than previously available techniques, permitting missing data and even informative censoring. There is also a heightened interest in using quality of life endpoints[208,212,214] and this gives rise to longitudinal data. In the cancer literature the fallacious approach of comparing survivals of responders versus non-responders has been criticized by statisticians[209,218,220] and this has clarified to many clinicians the importance of designing randomized studies large enough to directly address the effect of treatment on survival. It has also given rise to a need for measures to directly address the palliative effects of treatment. In many areas of clinical trials survival time is prolonged and valid short-term surrogate endpoints are needed. This has led to a re-evaluation of the requirements of surrogate endpoints and an increased emphasis on the identification of surrogates.[211,215,216,219,221]

7. REGRESSION MODELS

Regression models[223-234] are widely used in the analysis of clinical trials; the logistic model for binary response data and the proportional hazards model for right censored data are particularly popular. In the past decade a substantial amount of useful work has been devoted to the development of statistical tools for the diagnosis of model problems and evaluation of lack of fit of these models. There has long been confusion on the role of randomization in regression based estimates of treatment effects. Gail *et al.*,[225,226] Lagakos and Schoenfeld[228] and Morgan and Elashoff[229,230] published important papers on the effect of omitting covariates on the estimators of treatment effects in randomized clinical trials. Bayesians used to be fond of saying that randomization was of no importance in inference. The fallacy in their argument was that they assumed that their model, their likelihood function, was correct. This assusmes away much of the challenge of solving real-world problems with statistics. The importance of randomization in Bayesian inference was pointed out in an important paper by Rubin[232] and was more recently discussed by Smith and Sugden.[234] There had been a conceptual gap, however, in the use of randomization for treatment assignment and the use of regression modelling for data analysis. Work in the past decade has helped to bridge this gap. The process of regression modelling remains, however, difficult and hazardous. We have better tools for evaluating model adequacy, but we still do not know enough about how the process of model development influences the estimates of treatment effect. Some insight into this process has resulted from the investigations of

Schluchter and Forsythe[233] and Beach and Meier[224] on the effect of *post hoc* selection of covariates on the estimate of treatment effect. If the process of model development could be completely specified, then we could study this process through re-sampling methods (for example, Altman,[223] Picard[231]). Usually, however, the model development process contains subjective or poorly specified components. Whereas these concerns may prove to be less problematic in large samples than in small, it is an area that warrants further research.[227]

8. SURVIVAL ANALYSIS

The growth of clinical trials over the past 25 years has resulted in extensive research in the development of methods for the analysis of right censored data. In the past 10 years, this has included substantial effort on the decvelopment of useful diagnostic and goodness-of-fit tools for use in conjunction with the proportional hazards model[235-248] and the study of statistical significance tests for censored data.[249-275] Schoenfeld and Tsiatis,[268] for example, developed an interesting test for comparing survival distributions that is applicable to highly stratified data and retains power even if the strata are not highly prognostic. There was also research on estimation of survival distributions and hazard functions[276-292] including methods for estimating confidence intervals for the median and other quantiles of survival curves. This work is useful because median survival is often reported and is frequently a misleading summary statistic.

New univariate[293-302] and multi-variate[303-320] regression models for censored data have been introduced. These have included multi-state models that represent intermediate disease states such as remission and relapse prior to the ultimate absorbing state, death. Additional assumptions are entailed with use of the multi-state models and we need more experience with their use. They may require more data than conventional approaches, however, in order to have confidence in the model assumptions made.

The end of the decade also saw three papers discussing measures of explained variation for survival models.[321-323] Whereas 'statistically significant' covariates are common, most models of prognosis for patients in a single clinical trial explain little of the total variation in survival. This is sometimes not recognized.

9. PROBLEMS OF MULTIPLICITY

There are many problems of multiplicity in clinical trials: multiple endpoints, patient subsets, interim analyses and treatment contrasts.[324-347] Such problems represent some of the most difficult statistical challenges that we face and account for a large portion of misleading results in the medical literature. Our clinical trials are usually not designed large enough to answer a multitude of questions and our most commonly employed statistical methods are not appropriate for the way they are often used.[341]

Most clinical trials involve multiple endpoints. Often, there is interest in making separate inferences about each measure without assuming that treatment effects will be similar for all. An overall test of equality of the treatments with regard to all the endpoints may be useful but rejection of the global null hypothesis is not sufficiently informative. In such circumstances, separate inference, with the possible use of multiple comparison procedures is appropriate. A second setting is where the endpoints are alternative measures of the same fundamental quantity. In this case separate analyses may lead to highly confusing results and the most appropriate approach is often to treat the multiple endpoints as a single multivariate outcome. O'Brien[338] and Pocock *et al.*[340] studied methods of this type. There has been a tremendous amount of research on the analysis of multivariate outcomes that are of the repeated measurement or longitudinal data type. The methods have been developed for binary, normal and generalized linear models, random and fixed effects, non-informative and informative censoring.

The developments in this area are too extensive to review here, but represent a major accomplishment of clinical trial methodology for the 1980's. The proceedings of a workshop on longitudinal data analysis published in *Statistics in Medicine* provides an excellent entry to this literature.[222]

Because patient populations are almost always heterogeneous, subset analysis is often an important component of the report of major clinical trials. Most experienced medical statisticians regard subset analysis as 'hypothesis generating' with findings to be tested in other studies and such findings are notoriously unreliable due to the magnitude of the multiplicity problem. There is a need for better statistical tools and practices for conducting subset analyses. Two useful practices are pre-specification of a very few subset hypotheses in order to reduce the magnitude of the multiplicity problem, and the use of interaction tests rather than significance tests within subsets. One tool that was developed in the past decade was the test for 'qualitative' treatment by subset interaction.[333] A qualitative interaction means that one treatment is superior for some subsets of patients whereas the alternative treatment is superior for other subsets.[339] Qualitative interactions are also sometimes called reversals. They are important because the usual quantitative interactions may represent a simple dependence on the scale with which treatment effect is measured. Gail and Simon[333] developed a likelihood ratio test for qualitative interactions among two treatments and K disjoint subsets. Shuster and Van Eys[344] studied regions of treatment preference with regard to a continuous covariate. Other tools for subset analysis have been reviewed by Simon.[345,346] One such tool is empirical Bayes analysis. This model, recently discussed by Davis and Leffingwell,[339] and previously by Louis,[336] represents the observed treatment effect in a subset as the sum of the true subset effect plus a normal error with mean zero and known variance. The subset values are assumed to be independently and identically distributed from an unknown normal mixing distribution. The parameters of the mixing distribution are estimated from the data and the posterior normal distribution of each subset effect is then calculated. Donner[330] and Durrleman and Simon[331] have also described Bayesian models. Fleiss[332] has discussed treatment by institution interactions in multi-center clinical trials.

10. COMPUTER INTENSIVE METHODS

The past decade saw the development and increased utilization of computer intensive methods of analysis in clinical trials.[348-373] One area is the use of exact non-parametric statistical significance tests for evaluating treatment contrasts.[358,361,364-370,373] Using these methods one need no longer worry about the accuracy of large-sample approximations to the distribution of the test statistic. Computationally efficient exact methods are available for the usual two sample, k sample and contingency table based methods with or without stratification. Exact methods are also available for logistic regression analysis.

There has also been an extensive amount of research devoted to generalizing the linear functional employed in regression analysis. One approach was the development of additive models.[356,360,372] This replaced the linear terms by non-linear smoothers. This approach is computationally very intensive and not directly applicable to logistic or proportional hazards models. An alternative generalization is that of regression splines.[352,359] The regression spline approach is computationally much less intensive, does not require specialized software and is applicable to the common non-linear regression models used for clinical trials.

Recursive partitioning methods were developed in this decade and have started to receive increasing attention.[349-351,357,362,363,371] The monograph on Classification and Regression Trees (CART) by Breiman *et al.*[349] stimulated much interest. CART, as originally described, was not applicable to survival data but was generalized for such data by several individuals. CART and its modifications are well suited to the development of prognostic classifications. Because the

classifications are based on Boolean combinations of simple univariate functions, they tend to be intuitively appealing to clinicians. There is a great risk of overfitting the data when using high dimensional multivariate methods. CART tries to avoid this by using cross-validation for model selection. This is an important aspect of CART and is more widely applicable. We are still learning about the relative merits of recursive partitioning and additive modelling (including the usual logistic and proportional hazard regression methods). CART is not, at this point, an inferential tool for analysing clinical trials. It is an exploratory tool for deriving predictions and classifications.

There are other areas of statistical computing that saw major advances in the decade of the 80's. These include the use of dynamic graphics for data analysis[348] and sample-reuse methods such as the bootstrap for estimating the sampling properties of estimators.[353-355]

11. REPORTS AND OVERVIEWS

The main purpose of clinical trials is to improve public health by generating and communicating knowledge about the presence or absence of treatment effectiveness. The body of knowledge on how to plan, conduct and analyse good clinical trials has improved substantially over the past two decades. But the impact of clinical trials on medical practice is limited by difficulties in the communication of results, the interpretation of reports by practising physicians and by the multiplicity of trials conducted. Shortcomings in the reporting of clinical trial results have been well documented in the past. The past decade has seen efforts to establish meaningful guidelines for the publication of clinical trial reports, efforts to get statisticians more involved in the medical refereeing process and further attempts to educate physicians concerning the effective conduct and reporting of clinical trials.[341,374-386]

In addition to difficulties with the content of individual clinical trial reports, the process of communication is distorted by publication bias.[387-393] Publication bias denotes the selection processes that determine whether a clinical trial will get reported in the literature, and if so, in which journal and with what level of publicity. Small clinical trials are less likely to be published if results are 'negative' either because the principal investigator is less motivated to prepare a manuscript or because a journal is less likely to accept it. This is less true of large randomized multi-centre clinical trials, but the journal in which the results will appear may be strongly influenced by whether the results are 'positive' or 'negative'. The journal in which a manuscript appears and the publicity associated with its publication may have a great influence on the impact of the clinical trial on medical practice. Recently the National Institutes of Health have been concerned that reports of important clinical trials be communicated rapidly and widely. The press conferences and mass mailings that have resulted can enhance the impact of results from such clinical trials.

Because of the large number of clinical trials being conducted, there is a serious risk that many of those that are statistically significant at conventional levels represent 'false positives'. This has been discussed by several authors.[156, 394-396] The probability of 'false positivity' for an aggregate of clinical trials depends strongly on the 'prevalence' of true clinically significant treatment differences as well as on the type I and II error rates as for diagnostic screening. When the prevalence or prior probability of positivity is low, then treatment effects that are statistically significant at the 0·05 level are less compelling. Consequently confirmatory trials are very important and interpreting clinical trials in isolation of 'related' trials is hazardous.

The issue of combining evidence from multiple clinical trials was prominent and controversial in the 1980's. Meta-analyses, or quantitative overviews, widely used in social sciences, have now been applied to several areas of medicine.[397-411] Meta-analysis generally results in a number

representing an average treatment effect and this has sometimes led to controversy over whether the clinical trials were similar enough with regard to treatments, populations, quality of conduct, etc. for meaningful combination. Meta-analysis attempts to circumvent the common practice of physicians of selecting one clinical trial to believe and rejecting the contradictory results of other non-identical but related trials. One can usually retrospectively rationalize why one variant of a treatment should work better than others. There is a problem, however, in averaging the results of clinical trials that are too diverse and meta-analysis can be misleading when employed in a mechanical manner without adequate knowledge of the disease and its treatment. To the extent that the studies are very similar, there is less controversy. The decade of the 80's saw contributions to the methodology of meta-analysis in medicine. Meta-analysis can be very useful and is likely to be more often employed in the future. If it is based on published clinical trial reports, however, its validity can be limited by publication bias and the poor quality of publications. Its success in estimating the effectiveness of a specific therapy also depends on the existence of large enough groups of patients properly randomized to the same treatments. Otherwise it gives only a rough answer to a rough question about the average effectiveness of a broad class of treatments for a broad class of patients. Whereas the answer to that question may be of some value, we usually expect to learn more from individual clinical trials. Meta-analysis can be useful but it is not a satisfactory alternative to large randomized clinical trials of specific treatments.

12. CONCLUSION

I apologize for the omission and cursory treatment of many important areas; my difficulty is a tribute to the vitality of the field. The growth of clinical trials has given us a broad base of talented and dedicated professional statisticians working in this field. These individuals are distributed in a wide variety of organizations including the pharmaceutical industry, universities and government. The 1990's will see many exciting medical developments but the need for statistical excellence in drug development and treatment evaluation will not decrease.

ACKNOWLEDGEMENT

I would like to express my appreciation to Ms. Leslie Dietrich for her excellent assistance in preparing this manuscript.

REFERENCES

Sequential designs

1. Armitage, P. *Sequential Medical Trials*, Blackwell Scientific Publications, 1960.
2. Armitage, P., Stratton, I. M. and Worthington, H. V. 'Repeated significance tests for clinical trials with a fixed number of patients and variable follow-up', *Biometrics*, **41**, 353–359 (1985).
3. DeMets, D. L. 'Practical aspects in data monitoring: A brief review', *Statistics in Medicine*, **6**, 743–760 (1987).
4. DeMets, D. and Gail, M. 'Use of logrank tests and group sequential methods at fixed calendar times', *Biometrics*, **41**, 1039–1044 (1985).
5. DeMets, D. and Lan, G. 'An overview of sequential methods and their application in clinical trials', *Communications in Statistics A*, **13**, 2315–2338 (1984).
6. DeMets, D. 'Stopping guidelines vs. stopping rules: A practitioner's point of view', *Communications in Statistics A*, **13**, 2395–2417 (1984).
7. Enas, G. G., Dornseif, B. E., Sampson, C. B., Rockhold, F. W. and Wuu, J. 'Monitoring versus interim analysis of clinical trials: A perspective from the pharmaceutical industry', *Controlled Clinical Trials*, **10**, 57–70 (1989).

8. Fairbanks, K. and Madsen, R. '*P* values for tests using a repeated significance test design', *Biometrika*, **69**, 69–74 (1982).
9. Fleming, T., Green, S. J. and Harrington, D. 'Considerations for monitoring and evaluating treatment effects in clinical trials', *Controlled Clinical Trials*, **5**, 55–66 (1984).
10. Fleming, T., Harrington, D. and O'Brien, P. O. 'Designs for group sequential tests', *Controlled Clinical Trials*, **5**, 348–361 (1984).
11. Geary, D. N. 'Sequential testing in clinical trials with repeated measurements', *Biometrika*, **75**, 311–318 (1988).
12. Geller, N. L. and Pocock, S. J. 'Interim analyses in randomized clinical trials: Ramification and guidelines for practitioners', *Biometrics*, **43**, 213–223 (1987).
13. Jennison, C. 'Efficient group sequential tests with unpredictable group sizes', *Biometrika*, **74**, 155–165 (1987).
14. Kim, K. and DeMets, D. 'Design and analysis of group sequential tests based on the type I error spending function', *Biometrika*, **74**, 149–154 (1987).
15. Lan, K. K. G. and DeMets, D. 'Changing frequency of interim analysis in sequential monitoring', *Biometrics*, **45**, 1017–1020 (1989).
16. Lan, K. K. G. and DeMets, D. 'Discrete sequential boundaries for clinical trials', *Biometrika*, **70**, 659–663 (1983).
17. McPherson, K. 'On choosing the number of interim analyses in clinical trials', *Statistics in Medicine*, **1**, 25–36 (1982).
18. Pocock, S. J. 'Group sequential methods in the design and analysis of clinical trials', *Biometrika*, **64**, 191–199 (1977).
19. Pocock, S. J. 'Interim analyses for randomized clinical trials: The group sequential approach', *Biometrics*, **38**, 153–162 (1982).
20. Rosner, G. and Tsiatis, A. A. 'The impact that group sequential tests would have made on ECOG clinical trials', *Statistics in Medicine*, **8**, 505–516 (1989).
21. Rubinstein, L. V. and Gail, M. H. 'Monitoring rules for stopping accrual in comparative survival studies', *Controlled Clinical Trials*, **3**, 325–343 (1982).
22. Sellke, T. and Siegmund, D. 'Sequential analysis of the proportional hazards model', *Biometrika*, **70**, 315–326 (1983).
23. Slud, E. V. 'Sequential linear rank tests for two-sample censored survival data', *Annals of Statistics*, **12**, 551–571 (1984).
24. Slud, E. and Wei, L. J. 'Two-sample repeated significance tests based on the modified Wilcoxon statistic', *Journal of the American Statistical Association*, **77**, 862–868 (1982).
25. Tang, D., Grecco, C. and Geller, N. L. 'Design of group sequential clinical trials with multiple endpoints', *Journal of the American Statistical Association*, **84**, 776–779 (1989).
26. Tsiatis, A. A., Rosner, G. L. and Tritchler, D. L. 'Group sequential tests with censored survival data adjusting for covariates', *Biometrika*, **72**, 365–373 (1985).
27. Tsiatis, A. A. 'The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time', *Biometrika*, **68**, 311–315 (1981).
28. Tsiatis, A. A. 'Repeated significance testing for a general class of statistics used in censored survival analysis', *Journal of the American Statistical Association*, **77**, 855–861 (1982).
29. Wang, S. and Tsiatis, A. A. 'Approximately optimal one-parameter boundaries for group sequential trials', *Biometrics*, **43**, 193–199 (1977).
30. Whitehead, J. and Stratton, I. 'Group sequential clinical trials with triangular continuation regions', *Biometrics*, **39**, 227–236 (1983).
31. Whitehead, J. *The Design and Analysis of Sequential Clinical Trials*, Ellis Harwood, Chichester, 1982.

Early rejection of experimental treatments

32. Canner, P. L. 'Monitoring of data for evidence of adverse or beneficial effects', *Controlled Clinical Trials*, **4**, 467–483 (1983).
33. Chi, P. Y., Bristol, D. and Castellana, J. V. 'A clinical trial with an interim analysis', *Statistics in Medicine*, **5**, 387–392 (1986).
34. Choi, S., Smith, P. and Becker, D. 'Early decision in clinical trials when the treatment differences are small: experience of a controlled trial in head trauma', *Controlled Clinical Trials*, **6**, 280–288 (1985).
35. DeMets, D. and Ware, J. 'Asymmetric group sequential boundaries for monitoring clinical trials', *Biometrika*, **69**, 661–663 (1982).

36. DeMets, D. and Ware, J. 'Group sequential methods for clinical trials with a one-sided hypothesis', *Biometrika*, **67**, 651–660 (1980).
37. Emerson, S. and Fleming, T. 'Symmetric group sequential test designs', *Biometrics*, **45**, 905–923 (1989).
38. Freedman, L. S., Lowe, D. and Macaskill, P. 'Stopping rules for clinical trials incorporating clinical opinion', *Biometrics*, **40**, 575–586 (1984).
39. Gould, A. and Pecore, V. 'Group sequential methods for clinical trials allowing early acceptance of H_0 and incorporating costs', *Biometrika*, **69**, 75–80 (1982).
40. Lan, K. K. G. and Friedman, L. 'Monitoring boundaries for adverse effects in long-term clinical trials', *Controlled Clinical Trials*, **7**, 1–7 (1986).

Stochastic curtailment

41. Anderson, P. K. 'Conditional power calculations as an aid in the decision whether to continue a clinical trial', *Controlled Clinical Trials*, **8**, 67–74 (1987).
42. Dambrosia, J. M. and Greenhouse, S. W. 'Early stopping for sequential restricted tests of binomial distributions', *Biometrics*, **39**, 695–710 (1983).
43. DeMets, D. and Halperin, M. 'Early stopping in the two-sample problem for bounded random variables', *Controlled Clinical Trials*, **3**, 1–11 (1982).
44. DeMets, D., Hardy, R., Friedman, L. and Lan, K. K. G. 'Statistical aspects of early termination in the beta-blocker heart attack trials', *Controlled Clinical Trials*, **5**, 362–372 (1984).
45. Halperin, M., Lan, K. K. G., Ware, J. H., Johnson, N. J. and DeMets, D. 'An aid to data monitoring in long term clinical trials', *Controlled Clinical Trials*, **3**, 311–324 (1982).
46. Lan, K. K. G., Simon, R. and Halperin, M. 'Stochastically curtailed tests in long-term clinical trials', *Communications in Statistics— Sequential Analysis*, **1**, 207–219 (1982).
47. Pasternack, B. 'A note on data monitoring, incomplete data and curtailed testing', *Controlled Clinical Trials*, **5**, 217–222 (1984).

Repeated confidence intervals

48. Durrleman, S. and Simon, R. 'Planning and monitoring of equivalence studies', *Biometrics*, **46** 329–336 (1990).
49. Jennison, C. and Turnbull, B. 'Repeated confidence intervals for the median survival time', *Biometrika*, **72**, 619–626 (1985).
50. Jennison, C. and Turnbull, B. 'Interim analysis: The repeated confidence interval approach (with discussion)', *Journal of the Royal Statistical Society Series B*, **51**, 305–362 (1989).
51. Jennison, C. and Turnbull, B. 'Repeated confidence intervals for group sequential clinical trials', *Controlled Clinical Trials*, **5**, 33–46 (1984).
52. Lai, T. L. 'Incorporating scientific, ethical and economic considerations into the design of clinical trials in the pharmaceutical industry: a sequential approach', *Communications in Statistics A*, **13**, 2355–2368 (1984).

Two-stage designs

53. Bristol, D. 'A one-sided interim analysis with binary outcomes', *Controlled Clinical Trials*, **9**, 206–211 (1988).
54. Case, L. D., Morgan, T. M. and Davis, C. E. 'Optimal restricted two-staged designs', *Controlled Clinical Trials*, **8**, 146–156 (1987).
55. Ellenberg, S. S. and Eisenberger, M. A. 'An efficient design for phase III studies of combination chemotherapies', *Cancer Treatment Reports*, **69**, 1147–1154 (1985).
56. Thall, P. F., Simon, R. M. and Ellenberg, S. S. 'A two-stage design for choosing among several experimental treatments and a control in clinical trials', *Biometrics*, **45**, 537–547 (1989).
57. Thall, P. F., Simon, R. M. and Ellenberg, S. S. 'Two stage section and testing designs for comparative clinical trials', *Biometrika*, **75**, 303–310 (1988).
58. Thall, P. F., Simon, R. M. and Ellenberg, S. S. 'Optimal two-stage designs for clinical trials with binary responses', *Statistics in Medicine*, **7**, 571–579 (1988).
59. Whitehead, J. 'Sample sizes for phase II and phase III clinical trials, an integrated approach', *Statistics in Medicine*, **5**, 459–464 (1986).
60. Wieand, S. and Therneau, T. 'A two-stage design for randomized trials with binary outcome', *Controlled Clinical Trials*, **8**, 20–28 (1987).

Bayesian methods

61. Achcar, J., Brookmeyer, R. and Hunter, W. 'An application of Bayesian analysis to medical follow-up data', *Statistics in Medicine*, **4**, 509–520 (1985).
62. Berry, D. A. 'Interim analysis in clinical trials: Classical vs. Bayesian approaches', *Statistics in Medicine*, **4**, 521–526 (1985).
63. Berry, D. A. 'Monitoring accumulating data in a clinical trial', *Biometrics*, **45**, 1197–1211 (1989).
64. Berry, D. A. 'Interim analyses in clinical trials: Classical vs. Bayesian approaches', *Statistics in Medicine*, **4**, 521–526 (1985).
65. Berry, D. A. and Ho, C. H. 'One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach', *Biometrics*, **44**, 219–227 (1988).
66. Berry, D. A. 'Interim analysis in clinical trials: The role of the likelihood principle', *The American Statistician*, **41**, 117–122 (1987).
67. Choi, S. and Pepple, P. 'Monitoring clinical trials based on predictive probability of significance', *Biometrics*, **45**, 317–324 (1989).
68. Freedman, L. S. and Spiegelhalter, D. J. 'The assessment of subjective opinion and its use in relation to stopping rules for clinical trials', *The Statistician*, **33**, 153–160 (1983).
69. Freedman, L. S. and Spiegelhalter, D. S. 'Comparison of Bayesian with group sequential methods for monitoring clinical trials', *Controlled Clinical Trials*, **10**, 357–367 (1989).
70. Lachin, J. M. 'Sequential clinical trials for normal variates using interval composite hypothesis', *Biometrics*, **37**, 87–101 (1981).
71. Racine, A., Grieve, A. P., Fluehler H., Smith A. F. M. 'Bayesian methods in practice: Experiences in the pharmaceutical industry', *Applied Statistics*, **35**, 121–140 (1986).
72. Spiegelhalter, D. S. and Freedman, L. S. 'A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion', *Statistics in Medicine*, **5**, 1–13 (1986).
73. Spiegelhalter, D. S. and Freedman, L. S. 'Bayesian approaches to clinical trials', *In* Bernardo, J. M., DeGroot, M. H., Lindly, D. V. and Smith, A. F. M. (eds), *Bayesian Statistics 3*, North-Holland, Amsterdam, 1988.
74. Spiegelhalter, D. S., Freedman, L. S. and Blackburn, P. K. 'Monitoring clinical trials: conditional or predictive power', *Controlled Clinical Trials*, **7**, 8–17 (1986).

Estimation after sequential stopping

75. Atkinson, E. N. and Brown, B. W. 'Confidence limits for probability of response in multistage phase II clinical trials', *Biometrics*, **41**, 741–744 (1985).
76. Chang, M. 'Confidence intervals for a normal mean following a group sequential test', *Biometrics*, **45**, 247–254 (1989).
77. Chang, M. N. and O'Brien, P. C. 'Confidence intervals following group sequential tests', *Controlled Clinical Trials*, **7**, 18–26 (1986).
78. Chang, M. N., Wieand, H. S. and Chang, V. T. 'The bias of the sample proportion following a group sequential phase 2 design', *Statistics in Medicine*, **8**, 563–570 (1989).
79. Duffy, D. E. and Santer, T. J. 'Confidence intervals for a binomial parameter based on multistage tests', *Biometrics*, **43**, 81–93 (1987).
80. Hughes, M. D. and Pocock, S. J. 'Stopping rules and estimation problems in clinical trials', *Statistics in Medicine*, **7**, 1231–1242 (1988).
81. Jennison, C. and Turnbull, B. W. 'Confidence intervals for a binomial parameter following a multistage test with application to MIL-STD 105D and medical trials', *Technometrics*, **25**, 49–58 (1983).
82. Kim, K. and DeMets, D. 'Confidence intervals following group sequential tests in clinical trials', *Biometrics*, **43**, 857–864 (1987).
83. Kim, K. 'Improved approximation for estimation following closed sequential tests', *Biometrika*, **75**, 121–128 (1988).
84. Kim, K. 'Point estimation following group sequential tests', *Biometrics*, **45**, 613–618 (1987).
85. Pocock, S. J. and Hughes, M. D. 'Practical problems in interim analyses, with particular regard to estimation', *Controlled Clinical Trials*, **10**, 209S–221S (1989).
86. Skovlund, E. and Walloe, L. 'Estimation of treatment difference following a sequential clinical trial', *Journal of the American Statistical Association*, **84**, 823–828 (1989).
87. Tsiatis, A. A., Rosner, G. L. and Mehta, C. R. 'Exact confidence intervals following a group sequential test', *Biometrics*, **40**, 797–803 (1984).

88. Whitehead, J. 'Supplementary analysis at the conclusion of a sequential clinical trial', *Biometrics*, **42**, 461–472 (1986).
89. Whitehead, J. 'On the bias of maximum likelihood estimation following a sequential test', *Biometrika*, **73**, 573–582 (1986).

Sample size

90. Bernstein, D. and Lagakos, S. W. 'Sample size and power determination for stratified clinical trials', *Journal of Statistical Computing Simulations*, **8**, 65–73 (1978).
91. Blackwelder, W. C. '"Proving the null hypothesis" in clinical trials', *Controlled Clinical Trials*, **3**, 345–353 (1982).
92. Blackwelder, W. C. and Chang, M. A. 'Sample size graphs for "proving the null hypothesis"', *Controlled Clinical Trials*, **5**, 97–105 (1984).
93. Bristol, D. 'Sample sizes for constructing CI's and testing hypotheses', *Statistics in Medicine*, **8**, 803–821 (1989).
94. Conner, R. J. 'Sample size for testing differences in proportions for the paired-sample design', *Biometrics*, **43**, 207–211 (1987).
95. Dixon, D. O. and Simon, R. M. 'Sample size considerations for studies comparing survival curves using historical controls', *Journal of Clinical Epidemiology*, **41**, 1209–1214 (1988).
96. Donner, A. 'Approaches to sample size estimation in the designing of clinical trials—A review', *Statistics in Medicine*, **3**, 199–214 (1984).
97. Duffy, S. W. 'Asymptotic and exact power for the McNemar test and its analogue with R controls per case', *Biometrics*, **40**, 1005–1015 (1984).
98. Durrleman, S. D. and Simon, R. 'Planning and monitoring equivalence studies', *Biometrics*, **46**, 329–336 (1990).
99. Emrich, L. J. 'Sample size for historical control studies', *Statistics in Medicine*, **8**, 153–160 (1989).
100. Farewell, V. T. 'The use of mixture models for the analysis of survival data with long-term survivors', *Biometrics*, **38**, 1041–1046 (1982).
101. Freedman, L. S. 'Tables of the number of patients required in clinical trials using the logrank test', *Statistics in Medicine*, **1**, 121–130 (1982).
102. Goldman, A. 'Survivorship analysis when cure is a possibility: A Monte Carlo Study', *Statistics in Medicine*, **3**, 153–164 (1984).
103. Goodlow, J. L., Stablein, D. M. and Carter, W. H. 'Sample size for cancer therapy survival studies based on response surface methods', *Journal of Statistical Computing and Simulation*, **28**, 31–41 (1987).
104. Gross, A. J., Hunt, H. H., Cantor, A. B. and Clark, B. C. 'Sample size determination in clinical trials with an emphasis on exponentially distributed responses', *Biometrics*, **43**, 875–884 (1987).
105. Halpern, J. and Brown, B. W. 'Designing clinical trials with arbitrary specification of survival functions and for the logrank or generalized Wilcoxon test', *Controlled Clinical Trials*, **8**, 177–189 (1987).
106. Halpern, J. and Brown, B. W. 'Cure rate models: power of the logrank and generalized Wilcoxon tests', *Statistics in Medicine*, **6**, 483–490 (1987).
107. Korn, E. L. 'Sample size tables for bounding small proportions', *Biometrics*, **42**, 213–216 (1986).
108. Lachin, J. M. 'Introduction to sample size determination and power analysis for clinical trials', *Controlled Clinical Trials*, **2**, 93–114 (1981).
109. Lachin, J. M. and Foulkes, M. A. 'Evaluation of sample size and power for analyses of survival with allowance for non uniform patient entry, losses to follow-up, non-compliance and stratification', *Biometrics*, **42**, 507–519 (1986).
110. Lakatos, E. 'Sample sizes based on the log-rank statistic in complex clinical trials', *Biometrics*, **44**, 229–242 (1988).
111. Lakatos, E. 'Sample size in clinical trials with time dependent rates of losses and non compliance', *Controlled Clinical Trials*, **7**, 189–199 (1986).
112. Lee, Y. J. 'Quick and simple approximations of sample sizes for comparing two independent binomial distributions: Different-sample-size case', *Biometrics*, **40**, 239–242 (1983).
113. Makuch, R. W. and Johnson, M. F. 'Some issues in the design and interpretation of "negative" clinical studies', *Archives of Internal Medicine*, **146**, 986–989 (1986).
114. Makuch, R. W. and Simon, R. M. 'Sample size requirements for evaluating a conservative therapy', *Cancer Treatment Reports*, **62**, 1037–1040 (1978).
115. Makuch, R. W. and Simon, R. M. 'Sample size considerations for non-randomized comparative studies', *Journal of Chronic Disease*, **33**, 175–181 (1980).

116. Makuch, R. W. and Simon, R. M. 'Sample size requirements for comparing time-to-failure among k treatment groups', *Journal of Chronic Disease*, **35**, 861–867 (1982).
117. McHugh, R. B. and Lee, C. T. 'Confidence estimation and the size of a clinical trial', *Controlled Clinical Trials*, **5**, 157–164 (1984).
118. Munŏz, A. and Rosner, B. 'Power and sample size for a collection of 2×2 tables', *Biometrics*, **40**, 995–1004 (1984).
119. Morgan, T. M. 'Nonparametric estimation of duration of accrual and total study length for clinical trials', *Biometrics*, **43**, 903–912 (1987).
120. Morgan, T. M. and Elashoff, R. W. 'Effect of categorizing a continuous covariate on the comparison of survival times', *Journal of the American Statistical Association*, **81**, 917–921 (1986).
121. Morgan, T. M. and Elashoff, R. W. 'Effect of censoring on adjusting for covariates in comparison of survival times', *Communications in Statistics, Part A*, **15**, 1837–1854 (1986).
122. Moussa, M. A. A. 'Planning the size of survival time clinical trials with allowance for stratification', *Statistics in Medicine*, **7**, 559–569 (1988).
123. Nam, J. 'Simple approximation for calculating sample sizes for detecting linear trend in proportions', *Biometrics*, **43**, 701–705 (1987).
124. Noether, G. 'Sample size determination for some common nonparametric tests', *Journal of the American Statistical Association*, **82**, 645–647 (1987).
125. Palta, M. and Amini, S. B. 'Considerations of covariates and stratification in sample size determination for survival time studies', *Journal of Chronic Disease*, **38**, 801–809 (1985).
126. Rubinstein, L. V. Gail, M. H. and Santner, T. J. 'Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation', *Journal of Chronic Disease*, **34**, 469–479 (1981).
127. Schoenfeld, D. A. 'Sample size formula for the proportional hazards regression model', *Biometrics*, **39**, 499–503 (1983).
128. Schoenfeld, D. A. and Richter, J. R. 'Nomograms for calculating the numbers of patients for a clinical trial with survival as an endpoint', *Biometrics*, **38**, 163–170 (1982).
129. Schumaker, M., Olschewski, M. and Schmoor, C. 'The impact of heterogeneity on the comparison of survival curves', *Statistics in Medicine*, **6**, 773–784 (1987).
130. Sieh, F. Y. 'A simple method of sample size calculation for unequal-sample-size designs that use the logrank or t-test', *Statistics in Medicine*, **6**, 577–582 (1987).
131. Sposto, R. and Sather, H. N. 'Determining the duration of comparative clinical trials while allowing for cure', *Journal of Chronic Diseases*, **38**, 683–690 (1985).
132. Sposto, R. and Krailo, M. 'Use of unequal allocation on survival trials', *Statistics in Medicine*, **6**, 119–126 (1987).
133. Spriet, A. and Beiler, D. 'When can "non significantly different" treatments be considered as equivalent', *British Journal of Clinical Pharmacology*, **7**, 623–624 (1979).
134. Suissa, S. and Shuster, J. J. 'Exact unconditional sample sizes for the 2×2 binomial trial', *Journal of the Royal Statistical Society, Series A*, **148**, 317–327 (1985).
135. Taulbee, J. D. and Symons, M. J. 'Sample size and duration for cohort studies of survival time with covariables', *Biometrics*, **39**, 351–360 (1983).
136. Ury, H. and Fleiss, J. 'On approximate sample sizes for comparing two independent proportions with the use of Yates' corrections', *Biometrics*, **36**, 347–352 (1980).
137. Wittes, J. and Wallenstein, S. 'The power of the Mantel–Haenszel test', *Journal of the American Statistical Association*, **82**, 1104–1109 (1987).
138. Wu, W., Fisher, M. and DeMets, D. 'Sample sizes for long-term medical trial with time-dependent dropout and event rates', *Controlled Clinical Trials*, **1**, 111–123 (1980).
139. Wu, M. C. 'Comparison of changes', *Controlled Clinical Trials*, **9**, 32–46 (1988).

Treatment assignment methods

Randomization and stratification

140. Atkinson, A. C. 'Optimum biased coin designs for sequential clinical trials with prognostic factors', *Biometrika*, **69**, 61–67 (1982).
141. Begg, C. and Iglewicz, B. 'A treatment allocation procedure for sequential clinical trials', *Biometrics*, **36**, 81–90 (1980).
142. Begg, C. B. and Kalish, L. A. 'Treatment allocation for nonlinear models in clinical trials: The logistic model', *Biometrics*, **40**, 409–420 (1984).

143. Birkett, N. J. 'Adaptive allocation in randomized controlled trials', *Controlled Clinical Trials*, **6**, 146–155 (1985).
144. Cox, D. R. 'A remark on randomization in clinical trials', *Utilizing Mathematics*, **21A**, 245–252 (1982).
145. Friedman, L. W., Simon, R., Verter, J., Wittes, J. and Wittes, R. W. (eds) 'Proceedings of the workshop on Evaluation of Therapy', *Statistics in Medicine*, **3**, 305–486 (1984).
146. Halpern, J. and Brown, B. W. 'Sequential treatment allocation procedure in clinical trials-with particular attention to the analysis of results for the biased coin design', *Statistics in Medicine*, **5**, 211–230 (1986).
147. Kalish, L. A. and Begg, C. B. 'Treatment allocation methods: A review', *Statistics in Medicine*, **4**, 129–144 (1985).
148. Kalish, L. A. and Harrington, D. 'Efficiency of balanced treatment allocation for survival analysis', *Biometrics*, **44**, 815–822 (1988).
149. Lachin, J. M., Matts, J. P. and Wei, L. J. 'Randomization in clinical trials: conclusions and recommendations', *Controlled Clinical Trials*, **9**, 365–374 (1988).
150. McHugh, R. and Matts, J. 'Post stratification in the randomized clinical trial', *Biometrics*, **39**, 217–226 (1983).
151. Palta, M. 'Investigating maximum power losses in survival studies with nonstratified randomization', *Biometrics*, **41**, 487–504 (1985).
152. Palta, M. 'Magnitude and likelihood of loss resulting form nonstratified randomization', *Statistics in Medicine*, **1**, 261–266 (1982).
153. Pocock, S. J. and Simon, R. 'Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial', *Biometrics*, **31**, 103–115 (1975).
154. Reed, J. V. and Wickham, E. A. 'Practical experience of minimization in clinical trials', *Pharmaceutical Medicine*, **3**, 349–359 (1988).
155. Senn, S. J. 'Covariate imbalance and random allocation in clinical trials', *Statistics in Medicine*, **8**, 467–476 (1989).
156. Simon, R. 'Randomization and research strategy', *Cancer Treatment Reports*, **66**, 1083–1087 (1982).
157. Simon, R. 'Restricted randomization designs in clinical trials', *Biometrics*, **35**, 503–512 (1979).
158. Smith, R. L. 'Sequential treatment allocation using biased coin designs', *Journal of the Royal Statistical Society, Series B*, **46**, 519–543 (1984).
159. Soares, J. F. and Wu, C. F. J. 'Optimality of random allocation design for the control of accidental bias in sequential experiments', *Journal of Statistical Planning Information*, **11**, 81–87 (1985).
160. Taves, D. R. 'Minimization: A new method of assigning patients to treatment and control groups', *Clinical Pharmacology and Therapeutics*, **15**, 443–453 (1974).
161. Wei, L. J., Smythe, R. T. and Smith, R. L. 'K treatment comparisons with restricted randomization rules in clinical trials', *Annals of Statistics*, **14**, 265–274 (1986).

Adaptive treatment assignment

162. Bather, J. A. 'On the allocation of treatments in sequential medical trials', *International Statistical Review*, **53**, 1–14 (1985).
163. Bather, J. A. 'Randomized allocations of treatments in sequential experiments (with discussion)', *Journal of the Royal Statistical Society, Series B*, **43**, 265–292 (1981).
164. Berry, D. A. and Fristedt, B. *Bandit Problems: Sequential Allocation of Experiments*. Chapman and Hall, NY, 1985.
165. Cornell, R. G., Landenberger, B. D. and Bartlett, R. H. 'Randomized play-the-winner clinical trials', *Communications in Statistics, A— Theory and Methods*, **15**, 159–178 (1986).
166. Eick, S. G. 'Gittins procedures for bandits with delayed responses', *Journal of the Royal Statistical Society, Series B*, **50**, 125–132 (1988).
167. Glazebrook, K. D. 'On randomized dynamic allocation indices for sequential design of experiments', *Journal of the Royal Statistical Society, Series B*, **42**, 342–346 (1980).
168. Simon, R. 'Adaptive allocation in clinical trials', *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, 23–25 (1989).
169. Simon, R. 'Adaptive treatment assignment methods and clinical trials', *Biometrics*, **33**, 743–749 (1977).
170. Ware, J. H. 'Investigating therapies of potentially great benefit: ECMO (with discussion)', *Statistical Science*, **4**, 298–340 (1989).
171. Wei, L. J. 'Exact two-sample permutation tests based on the randomized play the winner rule', *Biometrics*, **75**, 603–606 (1988).

Phase II clinical trials

172. Atkinson, E. N., Brown, B. W. and Herson, J. 'KSTAGE: an interactive computer program for designing phase II clinical trials using predictive probability', *Computers in Biomedical Research*, **15**, 220–227 (1982).
173. Chang, M. N., Therneau, T., Wieand, H. S. and Cha, S. S. 'Designs for group sequential phase II clinical trials', *Biometrics*, **43**, 865–874 (1987).
174. Fleming, T. R. 'One sample multiple testing procedures for phase II clinical trials', *Biometrics*, **38**, 143–151 (1982).
175. Herson, J. 'Predictive probability early termination plans for phase II clinical trials', *Biometrics*, **35**, 775–784 (1979).
176. Herson, J. and Carter, S. K. 'Calibrated phase II clinical trials in oncology', *Statistics in Medicine*, **5**, 441–447 (1986).
177. Schaid, D. J., Ingle J. N., Wieand, S. and Ahmann, D. L. 'A design for phase II testing of anticancer agents within a phase III clinical trial', *Controlled Clinical Trials*, **9**, 107–118 (1988).
178. Simon, R. 'How large should a phase II trials of a new drug be?', *Cancer Treatment Reports*, **71**, 1079–1085 (1987).
179. Simon, R. 'Optimal two-stage designs for phase II clinical trials', *Controlled Clinical Trials*, **10**, 1–10 (1989).
180. Simon, R, Wittes, R. E. and Ellenberg, S. S. 'Randomized phase II clinical trials'. *Cancer Treatment Reports*, **69**, 1375–1381 (1985).
181. Sylvester, R. J. 'A Bayesian approach to the design of phase II clinical trials', *Biometrics*, **44**, 823–836 (1988).
182. Thall, P. F. and Simon, R. 'Incorporating historical control data in planning phase II clinical trials', *Statistics in Medicine*, **9**, 215–228 (1990).
183. Whitehead, J. 'Designing phase II studies in the context of a program of clinical research', *Biometrics*, **41**, 373–384 (1985).

Dose ranging trials

184. Anbar, D. 'Stochastic approximation methods and their use in bioassay and phase I clinical trials', *Communications in Statistics, Theory and Methods*, **13**, 2451–2467 (1984).
185. Bolognese, J. A. 'A Monte-Carlo comparison of three up-and-down designs for dose ranging', *Controlled Clinical Trials*, **4**, 187–196 (1983).
186. Budde, M. and Bauer, P. 'Multiple test procedures in clinical dose finding studies', *Journal of the American Statistical Association*, **84**, 792–796 (1989).
187. Chuang, C. 'Analysis of a titration study', *Statistics in Medicine*, **6**, 583–590 (1987).
188. O'Quigley, J., Pepe, M. and Fisher, L. 'Continual re-assessment method: A practical design for phase I clinical trials in cancer', *Biometrics*, **46**, 33–48 (1990).
189. Rademaker, A. W. 'Sample sizes in phase I toxicity studies', *Proceedings of the Biopharmaceutical Section*, 137–141, American Statistical Association, Alexandria, VA, 1989.
190. Ruberg, S. J. 'Contrasts for identifying the minimum efective dose', *Journal of the American Statistical Association*, **84**, 816–822 (1989).
191. Shih, W. J. 'Prediction approaches to sequentially searching for an optimal dose', *Biometrics*, **45**, 623–638 (1989).
192. Shih, W. J., Gould, A. L. and Hwang, I. K. 'The analysis of titration studies in phase III clinical trials', *Statistics in Medicine*, **8**, 583–592 (1989).
193. Storer, B. E. 'Design and analysis of phase I clinical trials', *Biometrics*, **45**, 925–937 (1989).
194. Tukey, J., Ciminera, J. L. and Heyse, J. F. 'Testing the statistical certainty of a response to increasing doses of a drug', *Biometrics*, **41**, 295–301 (1985).
195. Turri, M. and Stein, G. 'The determination of practically useful doses of new drugs: some methodological considerations', *Statistics in Medicine*, **5**, 449–457 (1989).

Other design issues

Two-period crossover

196. Brown, B. W. 'The crossover experiment for clinical trials', *Biometrics*, **36**, 69–79 (1980).
197. Fleiss, J. L. 'A critique of recent research on the two-treatment crossover design', *Controlled Clinical Trials*, **10**, 237–243 (1989).

198. Matthews, J. N. S., 'Recent developments in crossover designs', *International Statistical Review*, **56**, 117–127 (1988).
199. Willan, A. R. and Pater, J. L. 'Carryover and the two-period crossover clinical trial', *Biometrics*, **42**, 593–599 (1986).

N of one

200. Barlow, D. H. and Herson, M. *Single Case Experimental Designs: Strategies for Studying Behaviour Change*, Pergamon Press, New York, 1984.
201. Guyatt, G. H., Sackett, D., Taylor, D. W., Chong, J., Roberts, R. and Pugsley, S. 'Determining optimal therapy-randomized trials in individual patients', *New England Journal of Medicine*, **314**, 889–892 (1986).
202. Kazdin, A. E. *Single-case Research Designs: Methods for Clinical and Applied Settings*, Oxford University Press, New York, 1982.

Two by two factorial

203. Brittain, E. and Wittes, J. 'Factorial designs in clinical trials: The effects of non-compliance and subadditivity', *Statistics in Medicine*, **8**, 161–171 (1989).
204. Byar, D. P. and Piantadosi, S. 'Factorial designs for randomized clinical trials', *Cancer Treatment Reports*, **69**, 1055–1064 (1985).
205. Peto, R. 'Clinical trial methodology', *Biomedicine*, **28** (special issue), 24–36 (1978).
206. Stampfer, M. J., Buring, J. E., Willett, W., Rosner, B., Eberlein, K. and Hennekens, C. H. 'The 2×2 factorial design: Its application to a randomized trial of aspirin and carotene in U.S. physicians', *Statistics in Medicine*, **4**, 111–116 (1985).

Response surface

207. Carter, W. H., Wampler, G. L. and Stablein, D. M. *Regression Analysis of Survival Data in Cancer Chemotherapy*, Marcel Dekker (publ.), New York, 1983.

Endpoints

208. Aaronson, N. K. 'Quality of life assessment in clinical trials: Methodologic issues', *Controlled Clinical Trials*, **10**, 1955–2085 (1989).
209. Anderson, J. R., Cain, K. C. and Gelber, R. D. 'Analysis of survival by tumor response', *Journal of Clinical Oncology*, **1**, 710–719 (1983).
210. DeLong, E. R., Vernon, W. B. and Bollinger, R. R. 'Sensitivity and specificity of a monitoring test', *Biometrics*, **41**, 947–958 (1985).
211. Ellenberg, S. S. and Hamilton, J. M. 'Surrogate endpoints in clinical trials: cancer', *Statistics in Medicine*, **8**, 405–414 (1989).
212. Fayers, P. M. and Jones, D. R. 'Measuring and analyzing quality of life in cancer clinical trials: A review', *Statistics in Medicine*, **2**, 429–446 (1983).
213. Gail, M. H. 'Evaluating serial cancer markers studies in patients at risk of recurrent disease', *Biometrics*, **37**, 67–78 (1981).
214. Gelber, R. D., Gelman, R. S. and Goldhirsch, A. 'A quality-of-life oriented endpoint for comparing therapies', *Biometrics*, **45**, 781–796 (1989).
215. Herson, J. 'The use of surrogate endpoints in clinical trials', *Statistics in Medicine*, **8**, 403–404 (1989).
216. Hillis, A. and Seigel, D. 'Surrogate endpoints in clinical trials: Ophthalmologic disorders', *Statistics in Medicine*, **8**, 427–430 (1989).
217. Kay, R. 'A Markov model for analyzing cancer markers and disease states in survival studies', *Biometrics*, **42**, 855–866 (1981).
218. Mantel, N. 'An uncontrolled clinical trial— treatment response or spontaneous improvement?', *Controlled Clinical Trials*, **3**, 364–370 (1982).
219. Prentice, R. L. 'Surrogate endpoints in clinical trials: Definition and operational criteria', *Statistics in Medicine*, **8**, 431–440 (1989).
220. Weiss, G. B., Bunce, H. and Hokanson, J. A. 'Comparing survival of responders and non-responders after treatment: a potential source of confusion in interpreting cancer clinical trials', *Controlled Clinical Trials*, **4**, 43–52 (1983).

221. Wittes, J., Lakatos, E. and Probstfield, J. 'Surrogate endpoints in clinical trials: Cardiovascular diseases', *Statistics in Medicine*, **8**, 415–426 (1989).
222. Wu, W. C., Wittes, J. T., Zucker, D. and Kusek, J. (eds.) 'Proceedings of "Methods for Longitudinal Data Analysis in Epidemiological and Clinical Studies"', *Statistics in Medicine*, **7**, 1–136 (1988).

Regression models

223. Altman, D. G. and Anderson, P. K. 'Bootstrap investigation of the stability of a Cox regression model', *Statistics in Medicine*, **8**, 771–785 (1987).
224. Beach, M. L. and Meier, P. 'Choosing covariates in the analysis of clinical trials', *Controlled Clinical Trials*, **10**, 161S–175S (1989).
225. Gail, M. H., Wieand, S. and Piantadosi, S. 'Biased estimates of treatment effects in randomized experiments with nonlinear regressions and omitted covariates', *Biometrika*, **71**, 431–444 (1984).
226. Gail, M. H., Tan, W. Y. and Piantadosi, S. 'Tests for no treatment effect in randomized clinical trials', *Biometrika*, **75**, 57–64 (1988).
227. Harrell, F. E., Lee, K. L., Matchar, D. B. and Reichert, T. A. 'Regression models for prognostic prediction: Advantages, problems and suggested solutions', *Cancer Treatment Reports*, **69**, 1071–1077 (1985).
228. Lagakos, S. W. and Schoenfeld, D. A. 'Properties of proportional-hazards score tests under misspecified regression models', *Biometrics*, **40**, 1037–1048 (1984).
229. Morgan, T. M. 'Omitting covariates from the proportional hazards model', *Biometrics*, **42**, 993–995 (1983).
230. Morgan, T. M. and Elashoff, R. M. 'Effect of covariate measurement error in randomized clinical trials', *Statistics in Medicine*, **6**, 31–41 (1987).
231. Picard, R. R. and Cook, R. D. 'Cross validation of regression models', *Journal of the American Statistical Association*, **79**, 575–583 (1984).
232. Rubin, D. B. 'Bayesian inference for causal effects: The role of randomization', *Annals of Statistics*, **6**, 34–58 (1978).
233. Schluchter, M. D. and Forsythe, A. B. 'Post-hoc selection of covariates in randomized experiments', *Communications in Statistics A*, **14**, 679–699 (1985).
234. Smith, T. M. F. and Sugden, R. A. 'Sampling and assignment mechanisms in experiments, surveys and observational studies', *International Statistical Review*, **56**, 165–180 (1988).

Survival analysis

Proportional hazards model evaluation

235. Anderson, P. K. 'Testing goodness-of-fit of Cox's regression and life model', *Biometrics*, **38**, 67–77 (1982).
236. Arjas, E. 'A graphical method for assessing goodness of fit on Cox's proportional hazards model', *Journal of the American Statistical Association*, **83**, 204–212 (1988).
237. Cain, K. C. and Lange, N. T. 'Approximate case influence for the proportional hazards model', *Biometrics*, **40**, 493–500 (1984).
238. Ciampi, A. and Etezadi-Amoli, J. 'A general model for testing the proportional hazards and accelerated failure time hypothesis in the analysis of censored survival data with covariates', *Communications in Statistics A*, **14**, 651–667 (1985).
239. Gill, R. and Shumacher, M. 'A simple test of the proportional hazards assumption', *Biometrika*, **74**, 289–300 (1987).
240. Lakakos, S. W. 'The graphical evaluation of explanatory variables in proportional hazard regression models', *Biometrika*, **68**, 93–98 (1981).
241. Moreau, T., O'Quigley, J. and Lellouch, J. 'On Schoenfeld's approach for testing the proportional hazards assumption', *Biometrika*, **73**, 513–515 (1985).
242. Moreau, T., O'Quigley, J. and Mesbah, M. 'A global goodness-of-fit statistic for the proportional hazards model', *Applied Statistics*, **34**, 212–218 (1985).
243. Nagerlkerhe, N. J. D., Oosting, J. and Hard, A. A. M. 'A simple test of goodnes of fit of Cox's proportional hazards model', *Biometrics*, **40**, 483–486 (1984).
244. Reid, N. and Crepeau, H. 'Influence functions for proportional hazard regression', *Biometrika*, **72**, 1–10 (1985).

245. Schoenfeld, D. 'Chi-squared goodness-of-fit tests for the proportional hazards regression model', *Biometrika*, **67**, 145–154 (1980).
246. Schoenfeld, D. 'Partial residuals for the proportional hazard regression model', *Biometrika*, **69**, 239–241 (1982).
247. Storer, B. E. and Crowley, J. 'A diagnositic for Cox regression and general conditional likelihoods', *Journal of the American Statistical Association*, **80**, 139–147 (1985).
248. Wei, L. J. 'Testing goodness of fit for the proportional hazards model with censored data', *Journal of the American Statistical Association*, **79**, 649–652 (1984).

Significance testing

249. Beltanghady, M. S. and Frankowski, R. F. 'Effect of unequal censoring on size and power of logrank and Wilcoxon types of tests for survival data', *Statistics in Medicine*, **8**, 937–946 (1989).
250. Breslow, N. E., Edler, Ll. and Berger, J. 'A two-sample censored-data rank test for acceleration', *Biometrics*, **40**, 1049–1162 (1984).
251. Brookmeyer, R. and Crowley, J. 'A k-sample median test for censored data', *Journal of the American Statistical Association*, **77**, 433–440 (1982).
252. Brown, M. 'On the choice of variance for the logrank test', *Biometrika*, **71**, 65–74 (1984).
253. Crowley, J. and Johnson, R. (eds) *Survival Analysis*, IMS Lecture Notes, Monograph Series 2, Hayward; Institute of Mathematical Studies, 1982.
254. Cuzick, J. 'Rank tests for association with right censored data', *Biometrika*, **69**, 351–364 (1982).
255. Fleming, T. R., Harrington, D. P. and O'Sullivan, M. 'Supreme versions of the log-rank and generalized Wilcoxon statistics', *Journal of the American Statistical Association*, **82**, 312–320 (1987).
256. Fleming, T. R., O'Fallon, J. R., O'Brien, P. C. and Harrington, D. P. 'Modified Kolmogrov–Smirnov test procedures for right censored data', *Biometrics*, **36**, 607–626 (1980).
257. Freedman, L. S. and de Stavola, B. L. 'Comparing the effects of different durations of the same therapy', *Statistics in Medicine*, **7**, 1013–1022 (1988).
258. Gray, R. J. and Tsiatis, A. A. 'A linear rank test for use when the main interest is in differences in cure rates', *Biometrics*, **45**, 899–904 (1989).
259. Guilbaud, O. 'Exact Kolmogorov-type tests for left-truncated and/or right-censored data', *Journal of the American Statistical Association*, **83**, 213–221 (1988).
260. Halpern, M., Ware, J. H. and Wu, M. 'Conditional distribution-free tests for the two-sample problem in the presence of right censoring', *Journal of the American Statistical Association*, **75**, 638–645 (1980).
261. Harrington, D. P. and Fleming, T. R. 'A class of rank test procedures for censored survival data', *Biometrika*, **69**, 553–566 (1982).
262. Jones, M. P. and Crowley, J. 'A general class of nonparametric tests for survival analysis', *Biometrics*, **45**, 157–170 (1989).
263. Latta, R. 'A Monte Carlo study of some two-sample rank tests with censored data', *Journal of the American Statistical Association*, **76**, 713–719 (1981).
264. Mukerjee, H., Robertson, T. and Wright, E. T. 'Comparison of several treatments with a control using multiple contrasts', *Journal of the American Statistical Association*, **82**, 902–910 (1987).
265. O'Brien, P. C. and Fleming, T. R. 'A paired Prentice–Wilcoxon test for censored paired data', *Biometrics*, **43**, 169–180 (1987).
266. Pepe, M. S. and Fleming, T. R. 'Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data', *Biometrics*, **45**, 497–508 (1989).
267. Schoenfeld, D. 'The asymptotic properties of nonparametric tests for comparing survival distributions', *Biometrika*, **68**, 316–319 (1981).
268. Schoenfeld, D. A. and Tsiatis, A. A. 'A modified log rank test for highly stratified data', *Biometrika*, **74**, 167–176 (1987).
269. Stablein, D. M., Carter, W. H. and Novak, J. W. 'Analysis of survival data with nonproportional hazard functions', *Controlled Clinical Trials*, **2**, 149–159 (1981).
270. Stablein, D. M., Carter, W. H. and Wampler, G. L. 'Survival analysis of drug combinations using a hazards model with time-dependent covariates', *Biometrics*, **36**, 537–546 (1980).
271. Stablein, D. M. and Koutrouvelis, I. A. 'A two-sample test sensitive to crossing hazards in uncensored and singly censored data', *Biometrics*, **41**, 643–652 (1985).
272. Tarone, R. E. 'On the distribution of the maximum of the logrank and modified Wilcoxon statistics', *Biometrics*, **37**, 79–86 (1981).

273. Wei, L. J. 'A generalized Gehan and Gilbert test for paired observations that are subject to right censoring', *Journal of the American Statistical Association*, **75**, 634–637 (1980).
274. Woolson, R. F. and Lachenbruch, P. A. 'Rank tests for censored matched pairs', *Biometrika*, **67**, 597–606 (1980).
275. Woolson, R. F. and Lachenbruch, P. A. 'Rank tests for censored randomized block designs', *Biometrika*, **68**, 427–436 (1981).

Estimation and prediction

276. Afifi, A. A., Elashoff, R. M. and Lee, J. J. 'Simultaneous non-parametric confidence intervals for survival probabilities from censored data', *Statistics in Medicine*, **5**, 653–662 (1986).
277. Anderson, J. R., Bernstein, L. and Pike, M. C. 'Approximate confidence intervals for probabilities of survival and quantiles in life-table analysis', *Biometrics*, **38**, 407–416 (1982).
278. Anderson, J. R. and Bernstein, L. 'Asymptotically efficient two-step estimators of the hazard ratio for follow-up studies and survival data', *Biometrics*, **41**, 733–740 (1985).
279. Anderson, J. A. and Senthilselvan, A. 'Smooth estimates for the hazard function', *Journal of the Royal Statistical Society, Series B*, **42**, 322–327 (1980).
280. Bernstein, L., Anderson, J. and Pike, M. C. 'Estimation of the proportional hazard in two-treatment clinical trials', *Biometrics*, **37**, 513–520 (1981).
281. Brookmeyer, R. 'Prediction intervals for survival data', *Statistics in Medicine*, **2**, 485–495 (1983).
282. Brookmeyer, R. and Crowley, J. 'A confidence interval for the median survival time', *Biometrics*, **38**, 29–41 (1982).
283. Dorey, F. J. and Korn E. L. 'Effective sample sizes for confidence intervals for survival probabilities', *Statistics in Medicine*, **6**, 679–687 (1987).
284. Efron, B. 'Logistic regression, survival analysis and the Kaplan–Meier curve', *Journal of the American Statistical Association*, **83**, 414–425 (1988).
285. Emerson, J. D. 'Nonparametric confidence intervals for the median in the presence of right censoring', *Biometrics*, **38**, 17–27 (1982).
286. Korn, E. L. 'Censoring distributions as a measure of follow-up in survival analyses', *Statistics in Medicine*, **5**, 255–260 (1986).
287. Miller, R. G. 'What price Kaplan–Meier', *Biometrics*, **39**, 1077–1082 (1983).
288. Reid, N. 'Estimating the median survival time', *Biometrika*, **68**, 601–608 (1981).
289. Simon, R. and Lee, Y. J. 'Non-parametric confidence limits for survival probabilities and median survival time', *Cancer Treatment Reports*, **66**, 37–42 (1982).
290. Slud, E. V., Byar, D. P. and Green, S. B. 'A comparison of reflected versus test-based confidence intervals for the median survival time based on censored data', *Biometrics*, **40**, 587–600 (1984).
291. Slud, E. V. and Rubinstein, L. V. 'Dependent competing risks and summary survival curves', *Biometrika*, **70**, 643–649 (1983).
292. Thaler, H. T. 'Nonparametric estimation of the hazard ratio', *Journal of the American Statistical Association*, **79**, 290–293 (1984).

Univariate regression models

293. Aalen, O. O. 'Heterogeneity in survival analysis', *Statistics in Medicine*, **7**, 1121–1137 (1988).
294. Aalen, O. O. 'A linear regression model for the analysis of lifetimes', *Statistics in Medicine*, **8**, 907–926 (1989).
295. Etezadi-Amoli, J. and Ciampi, A. 'Extended hazard regression for censored survival data with covariates: A spline approximation for baseline hazard function', *Biometrics*, **43**, 181–192 (1987).
296. Finkelstein, D. M. and Wolfe, R. A. 'A semiparametric model for regression analysis of interval-censored failure time data', *Biometrics*, **41**, 933–945 (1985).
297. Finkelstein, D. M. 'A proportional hazards model for interval-censored failure time data', *Biometrics*, **42**, 845–854 (1986).
298. Hougaard, P. 'Life table methods for heterogeneous populations: Distributions describing the heterogeneity', *Biometrika*, **71**, 75–83 (1984).
299. Hougaard, P. 'Survival models for heterogeneous populations derived from stable distributions', *Biometrika*, **73**, 387–396 (1986).
300. Louis, T. 'Nonparametric analysis of an accelerated failure time model', *Biometrika*, **68**, 381–390 (1981).
301. Miller, R. G. and Halpern, J. 'Regression with censored data', *Biometrika*, **69**, 521–531 (1982).

302. Tibshirani, R. J. and Ciampi, A. 'A family of proportional and additive hazards models for survival data', *Biometrics*, **39**, 141–148 (1983).

Multivariate survival models

303. Aalen, O. O., Borgan, O., Keiding, N. and Thormann, J. 'Interaction between life history events. Nonparametri analysis for prospective and retrospective data in the presence of censoring', *Scandinavian Journal of Statistics*, **7**, 161–181 (1980).
304. Anderson, P. K. 'Miltistate models in survival analysis', *Statistics in Medicine*, **7**, 661–670 (1988).
305. Clayton, D. G. and Cuzick, J. 'Multivariate generalizations of the proportional hazards model (with discussion)', *Journal of the Royal Statistical Society, Series A*, **148**, 82–117 (1985).
306. Clayton, D. 'The analysis of event history data: A review of progress and outstanding problems', *Statistics in Medicine*, **7**, 819–842 (1988).
307. Cuzick, J. 'A review of semi-parametric models for life histories', *IMA Journal of Mathematics Applied in Medicine and Biology*, **1**, 323–332 (1984).
308. Dinse, G. E. and Larson, M. G. 'A note on semi-Markov models for partially censored data', *Biometrika*, **73**, 379–386 (1986).
309. Hsieh, F. Y., Crowley, J. and Tormey D. C. 'Some test statistics for use in multistate survival analysis', *Biometrika*, **70**, 111–119 (1983).
310. Huster, W. J., Brookmeyer, R. and Self, S. G. 'Modeling paired survival data with covariates', *Biometrics*, **45**, 145–156 (1989).
311. Jais, J. P. and Chastang, C. 'A transient state model for analyzing ventilatory status in a random clinical trial', *Statistics in Medicine*, **6**, 795–803 (1987).
312. Kay, R. 'Multistate survival analysis: an application in breast cancer', *Methods of Information in Medicine*, **23**, 157–162 (1984).
313. Matthews, D. E. 'Some observation on semi-Markov models for partially censored data', *Canadian Journal of Statistics*, **12**, 201–205 (1984).
314. Oakes, D. R. 'A model for association in bivariate survival data', *Journal of the Royal Statistical Society, Series B*, **44**, 414–428 (1982).
315. Oakes, D. R. 'Semiparametric inference in a model for association in bivariate survival data', *Biometrika*, **73**, 353–361 (1986).
316. Oakes, D. 'Bivariate survival models induced by frailties', *Journal of the American Statistical Association*, **84**, 487–493 (1989).
317. Rosner, B. 'Multivariate methods in ophthalmology with applications to other paired data situations', *Biometrics*, **40**, 1025–1036 (1984).
318. Seeber, G. U. H. 'On the regression analysis of tumour recurrence rates', *Statistics in Medicine*, **38**, 1363–1369 (1989).
319. Simon, R. and Makuch, R. W. 'A non-parametric graphical representation of the relationship between survival and the occurrence of an event: Application to responder versus non-responder bias', *Statistics in Medicine*, **3**, 35–44 (1984).
320. Voelkel, J. G. and Crowley, J. 'Nonparametric inference for a class of semi-Markov processes with censored observations', *The Annals of Statistics*, **12**, 142–160 (1984).

Measure of explained variation

321. Kent, J. T. and O'Quigley, J. 'Measures of dependence for censored survival data', *Biometrika*, **75**, 525–534 (1988).
322. Korn, E. L. and Simon, R. 'Measures of explained variation for survival data', *Statistics in Medicine*, **9**, 487–504 (1990).
323. Schemper, M. 'The explained variation in proportional hazards regression', *Biometrika*, **77**, 216–218 (1990).

Multiplicity

324. Bristol, D. R. 'Designing clinical trials for two-sided multiple comparisons with a control', *Controlled Clinical Trials*, **10**, 142–152 (1989).
325. Buyse, M. E. 'Analysis of clinical trial outcomes: Some comments on subgroup analyses', *Controlled Clinical Trials*, **10**, 1875–1945 (1989).

326. Byar, D. P. 'Assessing apparent treatment–covariate interactions in randomized clinical trials', *Statistics in Medicine*, **4**, 255–263 (1985).
327. Canner, P. L. 'Further aspects of data analysis', *Controled Clinical Trials*, **4**, 485–503 (1983).
328. Cox, D. R. 'Interaction', *International Statistical Review*, **52**, 1–32 (1984).
329. Davis, C. E. and Leffingwell, D. P. 'Empirical Bayesian estimates of subgroup effects in clinical trials', *Controled Clinical Trials*, **11**, 37–42 (1990).
330. Donner, A. A. 'A Bayesian approach to the interpretations of subgroup results in clinical trials', *Journal of Chronic Disease*, **35**, 429–435 (1982).
331. Durrleman, S. and Simon, R. 'The effect of diltiazem on mortality and reinfarction after myocardial infarction', *New England Journal of Medicine*, **320**, 123 (1989).
332. Fleiss, J. L. 'Analysis of data from multiclinic trials', *Controlled Clinical Trials*, **7**, 267–275 (1986).
333. Gail, M. and Simon, R. 'Testing for qualitative interactions between treatment effects and patients subsets', *Biometrics*, **41**, 361–372 (1985).
334. Gibson, J. M. and Overall, J. E. 'The superiority of a drug combination over each of its components', *Statistics in Medicine*, **8**, 1479–1484 (1989).
335. Laska, E. M. and Meisner, M. J. 'Testing whether an identified treatment is best', *Biometrics*, **45**, 1139–1151 (1989).
336. Louis, T. A. 'Estimating a population of parameter values using Bayes and empirical Bayes methods', *Journal of the American Statistical Association*, **79**, 393–398 (1984).
337. Mantel, H., Marocchi, A., Brambilla, P. and Baretta, R. 'Stratified analysis of multivariate clinical data. Application of a Mantel–Haenszel approach', *Statistics in Medicine*, **2**, 259–266 (1983).
338. O'Brien, P. 'Procedures for comparing samples with multiple endpoints', *Biometrics*, **40**, 1079–1088 (1984).
339. Peto, R. 'Statistical aspects of cancer trials', *in* Halnan, K. E. (ed.), *Treatment of Cancer*, Chapman and Hall, London, pp. 867–871, 1982.
340. Pocock, S. J., Geller, N. L. and Tsiatis, A. A. 'The analysis of multiple endpoints in clinical trials', *Biometrics*, **43**, 487–498 (1987).
341. Pocock, S. J., Hughes, M. D. and Lee, R. J. 'Statistical problems in the reporting of clinical trials', *New England Journal of Medicine*, **317**, 426–432 (1987).
342. Rosner, B. 'Multivariate methods in opthalmology with application to other paired data situations', *Biometrics*, **40**, 1025–1036 (1984).
343. Schneider, B. 'Analysis of clinical trial outcomes. Alternative approaches to subgroup analysis', *Controlled Clinical Trials*, **10**, 1765–1865 (1989).
344. Shuster, J. and Van Eys, J. 'Interaction between prognostic factors and tereatment', *Controlled Clinical Trials*, **4**, 209–214 (1983).
345. Simon, R. 'Statistical tools for subset analysis in clinical trials', in *Recent Results in Cancer Research*, **111**, 55–66, Springer-Verlag, Berlin-Heidelberg, 1988.
346. Simon, R. 'Patient subsets and variation in therapeutic efficacy', *British Journal of Clinical Pharmacology*, **14**, 473–482 (1982).
347. Snapinn, S. M. 'Evaluating the efficacy of a combination therapy', *Statistics in Medicine*, **6**, 657–665 (1987).

Computer intensive methods

348. Becker, R. A., Cleveland, W. S. and Wilks, A. R. 'Dynamic graphics for data analysis', *Statistical Science*, **2**, 355–395 (1987).
349. Breiman, L, Friedman, J. H. Olshen, R. A. and Stone, C. J. *Classification and Regression Trees*, Wadsworth International Group, Belmont, CA, 1984.
350. Ciampi, A., Lawless, L., McKinney, S. M. and Singhal, K. 'Regression and recursive partition strategies in the analysis of medical survival data', *Journal of Clinical Epidemiology*, **41**, 737–748 (1988).
351. Davis, R. B. and Anderson, J. R. 'Exponential survival trees', *Statistics in Medicine*, **8**, 947–962 (1987).
352. Durrleman, S. and Simon, R. 'Flexible regression models with cubic splines', *Statistics in Medicine*, **8**, 551–561 (1989).
353. Efron, B. and Tibshirani, R. 'Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy', *Statistical Science*, **1**, 54–77 (1986).
354. Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans*, Society of Industrial and Applied Mathematics, Philadelphia, PA, 1982.

355. Efron, B. 'Censored data and the bootstrap', *Journal of the American Statistical Association*, **76**, 312–319 (1981).
356. Friedman, J. H. and Silverman, B. W. 'Flexible parsimonious smoothing and additive modeling (with discussion)', *Technometrics*, **31**, 3–40 (1989).
357. Gordon, L. and Olshen, R. A. 'Tree-structured survival analysis', *Cancer Treatment Reports*, **69**, 1071–1078 (1985).
358. Haber, M. 'An exact unconditional test for the 2×2 comparative trial', *Psychological Bulletin*, **99**, 129–132 (1986).
359. Harrell, F. E., Lee, K. L. and Pollack, B. G. 'Regression models in clinical studies: Determining relationships between predictors and response', *Journal of the National Cancer Institute*, **80**, 1198–1202 (1989).
360. Hastie, T. and Tibshirani, R. 'Generalized additive models', *Statistical Science*, **1**, 297–318 (1986).
361. Hirji, K. F., Mehta, C. R. and Patel, N. R. 'Computing distributions for exact logistic regression', *Journal of the American Statistical Association*, **82**, 1110–1117 (1987).
362. Marshall, R. J. 'Partitioning methods for classification and decision making in medicine', *Statistics in Medicine*, **5**, 517–526 (1986).
363. Marubini, E., Morabito, A, and Valsecchi, M. G. 'Prognostic factors and risk groups: Some results given by using an algorithm suitable for censored survival data', *Statistics in Medicine*, **2**, 295–303 (1983).
364. Mehta, C. R., Patel, N. R. and Wei, L. J. 'Computing exact significance tests with restricted randomization rules', *Biometrika*, **75**, 295–302 (1988).
365. Mehta, C. R., Patel, N. R. and Tsiatis, A. A. 'Exact significance testing to establish treatment equivalence for ordered categorical data', *Biometrics*, **40**, 819–825 (1983).
366. Mehta, C. R. and Patel, N. R. 'A hybrid algorithm for Fisher's exact test on unordered $r \times c$ contingency tables', *Communications in Statistics*, **15**, 387–403 (1986).
367. Mehta, C. R. and Patel, N. R. 'A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables', *Journal of the American Statistical Association*, **78**, 427–434 (1983).
368. Mehta, C. R., Patel, N. R. and Senchaudhuri, P. 'Importance sampling for estimating exact probabilities in permutational inference', *Journal of the American Statistical Association*, **83**, 999–1005 (1988).
369. Mehta, C. R., Patel, N. R. and Gray, R. 'On computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables', *Journal of the American Statistical Association*, **80**, 969–973 (1985).
370. Pagano, M. and Halvorsen, K. T. 'An algorithm for finding the exact significance levels of $r \times c$ contingency tables', *Journal of the American Statistical Association*, **76**, 931–934 (1981).
371. Segal, M. R. and Bloch, D. A. 'A comparison of estimated proportional hazards and regression trees', *Statistics in Medicine*, **8**, 539–550 (1989).
372. Silverman, B. W. 'Some aspects of the spline smoothing approach to nonparameteric regression curve fitting (with discussion)', *Journal of the Royal Statistical Society, Series B*, **47**, 1–52 (1985).
373. Tritchler, D. 'An algorithm for exact logistic regression', *Journal of the American Statistical Association*, **79**, 709–711 (1984).

Reports and overviews

Reporting

374. Bailar, J. C. and Mosteller, F. 'Guidelines for statistical reporting in articles for medical journals: Amplifications and explanations', *Annals of Internal Medicine*, **108**, 266–273 (1988).
375. Chalmers, I. 'Proposal to outlaw the term "negative trials"', *British Medical Journal*, **290**, 1002 (1985).
376. Chalmers, T. C., Smith, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D. and Ambroz, A. 'A method for assessing the quality of a randomized control trials', *Controlled Clinical Trials*, **2**, 31–49 (1981).
377. Der Simonian, R., Charlette, L. J., McPeek, B. and Mosteller, F. 'Reporting on methods in clinical trials', *New England Journal of Medicine*, **306**, 1332–1336 (1982).
378. Ellenberg, S. S., Simon, R. and Wittes, R. E. (eds). 'Symposium on Methodology and Quality Assurance in Cancer Clinical Trials', *Cancer Treatment Reports*, **69**, 1039–1229 (1985).
379. Gardner, M. J. and Altman, D. G. 'Confidence intervals rather than p-values: estimation rather than hypothesis testing', *British Journal of Medicine*, **292**, 746–750 (1986).

380. George, S. L. 'Statistics in medical journals: A survey of current policies and proposals for editors', *Medical and Pediatric Oncology*, **13**, 109–112 (1985).
381. Kochar, M. S. 'The peer review of manuscripts— in need for improvement', *Journal of Chronic Disease*, **39**, 147–149 (1989).
382. Maxwell, C. 'Clinical trials, reviews and the journal of negative results', *British Journal of Clinical Pharmacology*, **1**, 15–18 (1981).
383. Mosteller, F., Gilberg, J. P. and McPeek, B. 'Reporting standards and research strategies for controlled clinical trials; agenda for the editors', *Controlled Clinical Trials*, **1**, 37–58 (1980).
384. Simon, R. and Wittes, R. E. 'Methodologic guidelines for reports of clinical trials', *Cancer Treatment Reports*, **69**, 1–3 (1985).
385. Simon, R. 'Confidence intervals for reporting results of clinical trials', *Annals of Internal Medicine*, **105**, 429–435 (1986).
386. Zelen, M. 'Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors', *Journal of Clinical Oncology*, **1**, 164–169 (1983).

Publication bias

387. Begg, C. B. and Berlin, J. A. 'Publication bias and dissemination of clinical research', *Journal of the National Cancer Institute*, **81**, 107–115 (1989).
388. Begg, C. B. 'A measure to aid in the interpretation of published clinical trials', *Statistics in Medicine*, **4**, 1–9 (1985).
389. Begg, C. B. and Berlin, J. A. 'Publication bias: A problem in interpreting medical data (with discussion)', *Journal of the Royal Statistical Society, Series A*, **151**, 419–463 (1989).
390. Berlin, J. A., Begg, C. B. and Louis, T. A. 'An assessment of publication bias using sample of published clinical trials', *Journal of the American Statistical Association*, **84**, 381–392 (1989).
391. Dickerson, K., Chan, S., Chalmers, T. C., Sacks, H. S. and Smith, H., Jr. 'Publication bias and clinical trials', *Controlled Clinical Trials*, **8**, 343–353 (1987).
392. Newcombe, R. G. 'Towards a reduction in publication bias', *British Medical Journal*, **295**, 656–659 (1987).
393. Simes, R. J. 'Confronting publication bias: a cohort design for meta-analysis', *Statistics in Medicine*, **6**, 11–30 (1986).

Study multiplicity

394. Ciampi, A. and Till, J. E. 'Null results in clinical trials. The need for a decision theory approach', *British Journal of Cancer*, **41**, 618–629 (1980).
395. Pater, J. L. and Willan, A. R. 'Clinical trials as diagnostic tests', *Controlled Clinical Trials*, **5**, 107–114 (1984).
396. Staquet, M. J., Rozenweig, M., Von Hoff, D. D. and Muggia, F. M. 'The delta and epsilon errors in the assessment of cancer clinical trials', *Cancer Treatment Reports*, **63**, 1917–1921 (1979).

Meta analysis

397. Berlin, J. A., Laird, N. M., Sacks, H. S. and Chalmers, T. T. 'A comparison of statistical methods for combining event rates from clinical trials', *Statistics in Medicine*, **8**, 141–151 (1989).
398. Boissel, J. P., Blanchard, J., Panak, E., Peyrieux, J. C. and Sacks, H. 'Consideration for the meta-analysis of randomized clinical trials: Summary of a panel discussion', *Controlled Clinical Trials*, **10**, 254–281 (1989).
399. Buyse, M. and Ryan, L. M. 'Issues of efficiency in combining proportions of deaths from several clinical trials', *Statistics in Medicine*, **6**, 565–576 (1987).
400. Chalmers, T. C., Berrier, J. Sacks, H. S., Levin, H., Reitman, D. and Nagalingam, R. 'Meta analysis of clinical trials as a scientific discipline II', *Statistics in Medicine*, **6**, 733–744 (1987).
401. Chalmers, T. C. 'Combinations of data from randomized control trials', *Biometrics*, **38**, 190–193 (1982).
402. Der Simonian, R. and Laird, N. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1988).
403. Goodman, S. N. 'Meta analysis and evidence', *Controlled Clinical Trials*, **10**, 188–204 (1989).
404. Hedges, L. and Olken, I. *Statistical Methods for Meta-Analysis*, Academic Press, Orlando, FL, 1987.

405. Iyengar, S. and Greenhouse, J. B. 'Selection models and the file drawer problem', *Statistical Science*, **3**, 109–135 (1988).
406. Light, R. and Pillemer, D. *Summing Up: The Science of Reviewing Research*, Harvard University Press, Cambridge, MA, 1984.
407. Meinert, C. L. 'Meta-analysis: Science or religion?', *Controlled Clinical Trials*, **10**, 257S–263S (1989).
408. Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A. and Chalmers, T. C. 'Meta-analyses of randomized controlled trials', *New England Journal of Medicine*, **316**, 450–455 (1987).
409. Simon, R. 'Overview of randomized clinical trials', *Cancer Treatment Reports*, **71**, 3–5 (1987).
410. Yusuf, S., Simon, R. and Ellenberg, S. (eds) 'Proceedings of the workshop on Methodologic Issues in Overview of Randomized Clinical Trials', *Statistics in Medicine*, **6**, 217–409 (1987).
411. Yusuf, S., Peto, R., Lewis, J., Collins, R. and Sleight, P. 'Beta blockade during and after myocardial infarction: an overview of randomized trials', *Progress in Cardiovascular Diseases*, **27**, 335–371 (1985).