PERGAMON

European Journal of Cancer 38 (2002) S162–S168

European Journal of Cancer

www.ejconline.com

Statistical methodology of phase III cancer clinical trials: advances and future perspectives

R. Sylvester*, M. Van Glabbeke, L. Collette, S. Suciu, B. Baron, C. Legrand, T. Gorlia, G. Collins, C. Coens, L. Declerck, P. Therasse

EORTC Data Center, 83 avenue E Mounier, Bte 11, 1200 Brussels, Belgium

Abstract

The methodology for conducting cancer clinical trials has undergone enormous changes over the past 25–30 years since the EORTC Data Center was created. The purpose of this paper is to highlight and to provide a historical perspective for the main methodological concepts, both practical and theoretical, which form the basis for the design and analysis of phase III cancer clinical trials within the EORTC Data Center. Some statistical aspects of other associated topics such as quality of life, health economics, meta-analysis and treatment outcome will also be briefly discussed. Finally, some future perspectives and topics for further statistical methodological research will be presented in order to spur statisticians to meet the challenge of efficiently designing and analysing the clinical trials of tomorrow. © 2002 Published by Elsevier Science Ltd.

Keywords: Statistics; Clinical trial; Phase III; Design; Analysis; Randomisation; Endpoints; Sample size; Meta-analysis

1. Introduction

The recognition of the importance of biostatistics in the field of cancer clinical trials has greatly increased since the first randomised clinical trial organised by the US National Cancer Institute (NCI) in acute lymphocytic leukaemia in 1954 [1]. Since the launch of the EORTC Data Center in 1974, major advances have been made in the development of the statistical methodology used in the design and analysis of cancer clinical trials. This is due to the work of, and interaction between, the cooperative group statisticians involved in the practical, day-to-day conduct of multicentre trials and academic statisticians who have developed the more theoretical aspects.

A number of authorities have also been actively involved in the development of methodological and statistical guidelines. Within the European Union, the Committee for Proprietary Medicinal Products (CPMP) has issued a number of Notes for Guidance, Points to Consider and Concept Papers dealing with biostatistical methodology [2–5]. In 1997, the CPMP also adopted a Note for Guidance on the Evaluation of Anticancer Medicinal Products in Man that was updated in 2000 [6].

The International Conference on Harmonization (ICH), a joint effort of the USA, European Union and Japan, has developed a number of ICH Harmonised Tripartite Guidelines which includes:

ICH Topic E6: Guidelines for Good Clinical Practice [7]
ICH Topic E8: General Considerations for Clinical Trials [8]
ICH Topic E9: Statistical Principles for Clinical Trials [9]
ICH Topic E10: Choice of Control Group in Clinical Trials [10]

It is recognised that the testing of a new therapy is a long-term project involving different types of trials during its development. The first step in planning a new trial is to precisely define its objectives and to determine the type of study to be carried out: phase I, phase II, or phase III. This paper will review some of the basic biostatistical and methodological concepts involved in conduct of phase III cancer clinical trials.

2. Design of phase III trials

2.1. Objectives

After the maximum tolerated dose (MTD) and the dose limiting toxicity (DLT) have been established in

* Corresponding author. Tel.: +32-2-774-1613; fax: 32-2-772-3545.
E-mail address: rsy@eortc.be (R. Sylvester).

0959-8049/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0959-8049(01)00442-7

phase I trials [11] and the therapy has been found to have some predefined minimal amount of antitumour activity in phase II trials [12], its relative efficacy is assessed in a randomised phase III trial, either alone or in combination with other treatments.

The possible goals of a phase III trial are:

- to determine the effectiveness of a treatment relative to the (treated) natural history of the disease (superiority trial).
- to determine the effectiveness of the new treatment compared with the best current standard therapy (superiority trial).
- to determine if a new treatment is as effective as the standard therapy, but is associated with less severe toxicity or a better quality of life (equivalence or non inferiority study).

A distinction should be made between trials attempting to show a difference in therapeutic effect (superiority trials) and trials designed to show non-inferiority (one sided) or equivalence (two sided) [3,9]. It is important not to conclude non-inferiority or equivalence if a test for superiority does not yield a statistically significant difference. Absence of evidence is not evidence of absence [13,14].

Phase III clinical trials should be simple, large scale, multicentre and randomised with key endpoints in order to have the maximal impact on day-to-day clinical practice. They should be planned to reliably detect small to moderate, but clinically worthwhile treatment differences, rather than to identify major breakthroughs. While a large survival improvement may be demanded of an aggressive therapy for a fatal condition, it is self-defeating to hope for more than a small survival improvement with an adjuvant therapy in good prognosis patients.

Preference should be given to large simple studies which compare two treatments which are as different as possible. Trials with more than two treatment arms are generally less efficient since they require proportionately more patients, and are often more difficult to recruit because patients must agree to receive any of the possible treatments. One possible exception is the use of 2×2 factorial designs to simultaneously study two different questions. Crossover trials are generally to be avoided since the underlying statistical assumptions for carrying out such trials are almost never satisfied in practice.

2.2. Patient selection

The entry of patients should be guided by the uncertainty principle: only if there is substantial uncertainty over the best treatment should the patient be randomised.

Two diverging attitudes may be adopted in defining eligibility criteria: they can be very precisely and narrowly defined so that only a small fraction of the available patients are eligible, thus making the patient sample as homogeneous as possible. This may be desirable to precisely identify treatment benefits in a subgroup of patients with well-defined prognostic features.

At the other extreme, they can be left as broad as possible so that most available patients can be entered into the trial, thus resulting in a faster patient accrual. This reduces the total duration of the study and yields a sample of patients which is more closely representative of the total patient population. The results can therefore be more readily extrapolated to patients outside the trial.

Narrow eligibility criteria are preferable to broad ones only when there are good *a priori* reasons to believe that the treatment will be beneficial only in a subgroup of patients. In the more common situation where the treatment effect is likely to be similar in different subgroups of patients, the best strategy is to choose eligibility criteria that are as broad as possible.

2.3. Randomisation and stratification

Although the first properly randomised trial (in the field of pulmonary tuberculosis) was carried out in 1947 [15], for both practical and ethical reasons randomisation was still not universally accepted as the method of choice for comparative studies in the early 1970s [16]. However, both the US NCI [17] and the EORTC [18] were instrumental in promoting the use of randomisation and highlighted the potential biases in doing studies using non-randomised or historical controls.

In an effort to try to make it easier for clinicians to enter patients in randomised trials, the Randomized Consent Design was introduced in 1979 [19]. In both the single and double randomised consent designs, patients are randomised prior to asking for their informed consent. This controversial design has a number of ethical and scientific problems which led to its disuse [20].

Today, centralised treatment allocation by a random process is the method of choice in comparative studies [21]. The EORTC Data Center has always provided a central randomisation by telephone for all its studies. This was extended to a central randomisation by modem in 1988 and via the Internet in 1994.

The EORTC prospectively stratifies the randomisation by institution and for a small number of the most important prognostic factors in order to assure an approximately equal balance of patients and prognostic factors in the treatment groups. Originally, the Data Center used the static method of randomised blocks [22]. However, if there were only a few patients within an institution, balance within the institution and hence within the trial was not assured. In 1983, the EORTC

replaced this method by the minimisation technique [22,23], a dynamic method which allows one to stratify for a larger number of prognostic factors. Its goal is to ensure a balance within each level of each stratification factor separately, but not necessarily within all the possible combinations of the different factors.

2.4. Endpoints

The choice of primary endpoint depends on the type of treatment and type of trial carried out.

- Adjuvant trials: Adjuvant therapy is given to patients who are clinically disease-free after a potentially 'curative' primary treatment but for whom there is a substantial risk of recurrence. The goal of an adjuvant trial is generally to compare the duration of survival, disease free survival or disease-free interval in two or more treatment groups. Time to local recurrence may also be of interest, but is problematic in the presence of competing risks. While a separate analysis of death due to malignant disease may be carried out, the main survival analysis always includes deaths due to any cause.
- Advanced disease trials: Advanced disease includes all patients for whom local treatment is usually no longer curative. This includes recurrent and locally advanced disease in which the disease is still confined to the region of the primary tumour as opposed to metastatic disease where the disease has spread to distant sites. When studying the effects of a new agent in advanced disease, the CPMP [6] recommends the following endpoints: progression-free survival, overall survival, response rate and symptom control/quality of life.

Response to treatment, duration of complete response, and duration of overall response are now most commonly assessed in accordance with RECIST [12], which replaced the World Health Organization (WHO) criteria in 2000. The RECIST criteria, developed in the context of phase II trials, are based on the decrease in size of prospectively selected 'target lesions'.

Toxicity is graded according to standard scales based on objective parameters. Originally based on the WHO classification, the 'Common Toxicity Criteria' (CTC) is now the most widely used scale (http://ctep.info.nih.gov/CTC3/ctc.htm).

In order to assess the symptomatic effect of a treatment, an additional efficacy endpoint that may be used is symptom control, which is supported by quality of life data. In recent years, much interest has focused on integrating both quality of life endpoints [24] and economic evaluations [25] into randomised phase III trials.

2.4.1. Surrogate endpoints

In an effort to draw conclusions on treatment efficacy based on information that becomes available more quickly than that for the primary endpoint (often duration of survival), the use of surrogate endpoints/biomarkers has become a hot topic of research over the past several years [26]. In general, short-term endpoints such as changes in marker levels, for example prostate specific antigen (PSA) in prostate cancer, and the time to first recurrence have not been shown to be adequate surrogate endpoints for long-term results such as the duration of survival. It should be underlined that response to therapy by itself is not an appropriate surrogate for therapeutic benefit, but only an indicator of antitumour activity.

2.5. Sample size

Large randomised trials are required to have a high power to detect small, but medically important, differences and to reduce the risk of random error. Meta-analyses have been extremely useful in showing that many trials have been designed based on overly optimistic expected treatment differences, and have thus been under-powered to detect smaller, more plausible differences.

The power to detect differences in survival and disease-free interval depends on the number of events (deaths or recurrences) in each group rather than on the total number of patients entered [27]. Thus, for adjuvant trials in patients with a good prognosis, especially large sample sizes and long-term follow-up are required.

Using the primary endpoint of interest, the determination of the sample size depends on the following factors:

- A realistic prior estimate for the endpoint of interest in the control group.
- Realistic estimates of the size of the plausible treatment effect and/or the medically worthwhile treatment effect.
- The size of the type I error α (false-positive rate; $\leqslant 0.05$; generally two-sided except for non-inferiority trials where it is recommended to take a one sided $\alpha = 0.025$) and type II error β (false-negative rate; $\leqslant 0.20$). $1-\beta$ is called the power.
- Realistic estimates of the expected accrual rate and duration of patient entry. In general, the expected duration of patient entry should not exceed 5 years.
- The duration of follow-up after closing the trial to patient entry.

Based on this information, the total number of events (in case of a time to event endpoint) and patients needed can be calculated in order to ensure a high power of detecting the postulated treatment difference at a prespecified significance level (α). Enough patients must be entered so that the required number of events can be observed within a given follow-up period.

Under the assumption of no interaction, it may be possible to use a 2×2 factorial design to study two questions without further increasing the sample size. Non-inferiority trials and equivalence studies, however, require considerably more patients than superiority trials since the hypothesised difference in such trials should be no more than one-third to one-half of the treatment effect of the control group detected in previous superiority studies [4].

3. Analysis of phase III trials

The foundations of the analysis of censored time to event data were published already during the period from 1958 to 1972:

1. estimation via the Kaplan–Meier product limit estimate [28]
2. comparison using the logrank test [29] and
3. prognostic factor analysis based on the Cox proportional hazards regression model [30].

Since then, a tremendous amount of research on this topic has been done, with advances in both computer hardware and software facilitating the practical implementation of the theoretical work.

3.1. Inclusion/exclusion of patients from analyses

All randomised patients, including those retrospectively found to be ineligible, should be treated and followed in accordance with the protocol whenever possible.

For efficacy comparisons, patients should be analysed according to the 'intent to treat' principle: all conclusions are based on all randomised patients according to the treatment group assigned by randomisation. Additional supportive 'per protocol' analyses may be carried out, for example analyses which are restricted to the eligible patients or eligible patients who started their treatment, especially in non-inferiority or equivalence studies where an intent to treat analysis may dilute the size of the treatment effect [9]. Toxicity analyses are based on all randomised patients who started their assigned treatment.

Guidelines (CONSORT) have been published in an attempt to standardise the format for the reporting of clinical trials and to improve the quality of published reports [31,32].

3.2. Switching between superiority and non-inferiority or equivalence

In some cases, the results of a superiority trial may only suggest non-inferiority or the results of a non-inferiority trial may suggest superiority. A non-inferiority trial may be able to be interpreted as showing evidence of superiority if the lower two-sided 95% Confidence Interval for the difference lies entirely above zero. However, a superiority trial cannot generally provide evidence of non-inferiority unless a non-inferiority margin has been prospectively defined in the protocol and the trial has been carried out according to the strict standards of non-inferiority trials [3].

3.3. Univariate analyses

Time to an event should be estimated by the Kaplan–Meier technique [27,28] and compared using a test such as the logrank that takes into account all the available data [27,29]. For response or other binary or categorical data, exact tests or Chi-square tests are to be employed.

3.4. Multivariate analyses

Retrospective stratification or multivariate models may be used to adjust the treatment comparison for the possible effect of prognostic factors. If one or more stratification factors other than institution have been used at the time of randomisation, then it is appropriate to take these factors into account during the analysis. Otherwise, the unadjusted analysis should be the primary analysis, with any adjusted analyses being supportive in nature. In any case, the policy for adjusting for prognostic factors should be prespecified in the protocol as various methodological problems exist [9,33].

The Cox model for time to event analyses [27,30] and logistic regression for binary endpoints are often recommended; however, their underlying assumptions should always be verified. In order to take into account additional sources of heterogeneity, proportional hazard models with random effects (frailty models) have recently received considerable attention [34].

3.5. Subgroup analyses/multiple comparisons

Comparing treatments in subgroups of patients suffers from a number of problems:

1. The probability of a false-positive result increases as the number of subgroup comparisons increases. In order to correct for multiple comparisons, the significance level used for each

comparison must be more extreme than the usual overall 5% level. The Bonferroni adjustment, "overall significance level/number of comparisons", provides a conservative significance level.
2. The reduced sample size in the subgroups and the low power to test for interactions makes it difficult to detect true subgroup effects.

All subgroup analyses should be specified *a priori* in the protocol. Any other subgroup analyses should be considered to be exploratory or hypothesis generating and must be interpreted with extreme caution [5,9,35].

3.6. Competing risks

A competing risk is the occurrence of an event which

- precludes the event of interest (death due to unrelated causes is a competing risk for death due to malignant disease) or
- modifies the probability of occurrence of the event of interest (the reporting of a local recurrence is less likely after the occurrence of distant metastases)

It is not appropriate to simply censor the event of interest at the date of the competing risk and calculate the Kaplan–Meier estimate, but rather cumulative incidence estimates should be calculated [36]. Likewise, the logrank test may also be inappropriate when making treatment comparisons in the presence of competing risks.

3.7. Interim analyses

If interim analyses are to be carried out, the following information should be clearly stated in the protocol:

1. The intention to perform interim analyses, their number and the timing of analyses.
2. The statistical stopping guidelines which will be used, whether it be to prematurely close the trial to patient entry or to decide whether to publish the results of the trial before observing the required number of events.

While a trial is still open to patient entry, the results of interim analyses of EORTC trials are not presented to trial participants. The report is submitted to the EORTC Independent Data Monitoring Committee (IDMC) composed of four permanent members (three medical doctors and one statistician) and two external experts who are not participating in the trial.

Early stopping may either be due to a large difference or due to 'negative' results (futility). Stopping guidelines based on a flexible alpha-spending function with an O'Brien–Fleming boundary (or a boundary approximately half way between O'Brien–Fleming and Pocock) are generally to be recommended for efficacy endpoints while a Pocock boundary may be more appropriate for safety endpoints [37].

The IDMC will take into account not only the statistical stopping guidelines, but also all other available information, for example adverse events, quality of life and results from other similar trials in making their recommendation [38].

4. Quality of life

The EORTC Quality of Life Unit was founded in 1993. The analysis of longitudinal quality of life data to assess the effect of treatment over time on the patient's quality of life poses an important methodological challenge [39]. In the analysis and interpretation of the results, one of the main problems is to identity the reasons for and properly take into account the patterns of missing data during follow-up [40], a subject which has been an important topic of research at the EORTC.

5. Health economics

The EORTC Health Economics Unit was also set up in 1993. Numerous research topics dealing with the methodology and application of economic studies in cancer have received attention. The EORTC has been involved in research dealing with statistical methods for calculating incremental cost-effectiveness ratios and cost-effectiveness acceptability curves and for assessing alternative methods for estimating mean survival.

6. Meta-analysis

Probably the biggest flaw in most clinical trials is that too few patients have been entered and an insufficient number of events have been observed to have a high probability of detecting a medically plausible difference in treatment efficacy should it exist. Meta-analysis (overview) is the process of using formal statistical methods to combine together the quantitative results of separate, but similar, studies in order to:

1. increase the statistical power to detect differences in treatment efficacy;
2. increase the precision of the estimated treatment effect.

Since 1994, the EORTC Meta-Analysis Unit has been active in the fields of adult leukaemia, breast cancer, bladder cancer, head and neck cancer, soft-tissue sarcoma, and a number of methodological research projects have been carried out. While meta-analyses can play a very important role in the overall scientific assessment of a treatment's efficacy [41], they are not a panacea or a cure all. In particular, they should not be a replacement for large-scale randomised clinical trials.

7. Treatment outcome studies

Treatment outcome studies assess the heterogeneity of patient outcome based on factors other than treatment, for example the institution, country or region where they are treated, and whether this may be due to differences in their baseline characteristics or in the quality of the treatment or supportive care received [42]. In 1998, the EORTC started a treatment outcome project which has focused on the methodological and practical aspects of applying frailty models in such studies and on the influence of the number of centres and patients per centre on the reliability of the results.

8. Future perspectives

In recent years, there have been important advances in the design of phase I and phase II trials, and in the design of phase III non-inferiority studies. Flexible rules have been set up for doing interim analyses along with the possibility of early stopping for futility. The analysis of prognostic factors has received considerable attention as has the analysis of quality of life, health economics and surrogate endpoints. While research in all these fields will continue, the techniques for handling missing data in quality of life and health economic studies and for identifying and assessing surrogate endpoints are especially in their infancy and much work still needs to be done in these areas.

As treatments become more effective and more cures are achieved, classical tests such as the logrank may not be appropriate. Cure rate models and other models for dealing with non-proportional hazards will be required. Diagnostic test studies will also play a role as cheaper and less invasive alternative methods for disease detection and recurrence are sought, thus spurring more research in Receiver Operating Characteristic (ROC) curve methodology.

Clinical studies generally fall into one of three categories: phase I, phase II and phase III. However, with the ever-increasing testing and assessment of certain new classes of non-cytotoxic anticancer agents, their development may not need to go through the classical phase I, II, III sequence to prove their efficacy or assess their toxicity. Indeed, it is not even clear in many cases what the primary endpoint should be. Under such circumstances, certain phases may sometimes be omitted or combined. Some trials may be designed up front as phase I/II or phase II/III, consecutively addressing two different primary endpoints. Thus, the design of trials carried out with non-cytotoxics will be a topic of considerable research in the coming years [43].

In the presence of a large number of potential interesting therapeutic regimens, more efficient methods are required to get them through phase II/phase III testing as only a limited number of regimens can actually be tested in full-scale phase III trials where survival is the main endpoint. In this perspective, the effect of treatment on alternative endpoints such as progression-free survival or on various biomarkers may play an important role in the screening process. As it is recognised that very large sample sizes are required in phase III trials, innovative techniques must also be sought to carry out intergroup trials in the most efficient manner possible.

As the different types of anticancer treatment and their rationale evolve over the coming years, it will be a challenge to biostatisticians to develop the methodology to efficiently design, carry out and analyse the clinical cancer trial of tomorrow.

References

1. Frei E, Holland JF, Schneiderman MA, *et al*. A comparative study of two regimens of combination chemotherapy in acute leukemia. *Blood* 1958, **13**, 1126–1148.
2. CPMP Working Party on Efficacy of Medicinal Products. Note for guidance III/3630/92-EN, biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. *Stat Med* 1995, **14**, 1659–1682.
3. Committee for Proprietary Medicinal Products (CPMP). *Points to Consider on Switching Between Superiority and Non-Inferiority*. CPMP/EWP/482/99, 2000.
4. European Agency for the Evaluation of Medicinal Products. *Concept Paper on the Development of a Committee for Proprietary Medicinal Products (CPMP). Points to Consider on Biostatistical/Methodological Issues Arising from Recent CPMP Discussions on Licensing Applications: Choice of Delta*. CPMP/EWP/2158/99, 1999.
5. European Agency for the Evaluation of Medicinal Products. *Concept Paper on the Development of a Committee for Proprietary Medicinal Products (CPMP). Points to Consider on Biostatistical/Methodological Issues Arising from Recent CPMP Discussions on Licensing Applications: Adjustment for Multiplicity and Related Topics*. CPMP/EWP/908/99, 1999.
6. Committee for Proprietary Medicinal Products (CPMP). *Note for Guidance on Evaluation of Anticancer Medicinal Products in Man*. CPMP/EWP/205/95 rev. 1, 2000.
7. ICH Topic E6. *Guideline for Good Clinical Practice*. CPMP/ICH/135/95, 1997.
8. ICH Topic E8. *General Considerations for Clinical Trials*. CPMP/ICH/291/95, 1997.
9. ICH Topic E9. *Statistical Principles for Clinical Trials*. CPMP/ICH/363/96, 1996.

10. ICH Topic E10. *Choice of Control Group in Clinical Trials*. CPMP/ICH/364/96, 1996.
11. Eisenhauer EA, Dwyer PJ, Christian M, Humphrey JS. Phase I clinical trial design in cancer drug development. *J Clin Oncol* 2000, **18**, 684–692.
12. Therasse P, Arbuck SG, Eisenhauer EA, *et al*. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000, **92**, 205–216 (http://www3.oup.co.uk/jnci/extra/920205.pdf; http://www.nci.nih.gov/bip/RECIST.htm).
13. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *Br Med J* 1995, **311**, 485.
14. Blackwelder W. Proving the null hypothesis in clinical trials. *Controlled Clin Trials* 1982, **3**, 345–353.
15. Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis. *Br Med J* 1948, **ii**, 769–782.
16. Gehan EA, Freireich EJ. Non-randomized controls in cancer clinical trials. *New Engl J Med* 1974, **290**, 198–203.
17. Byar DP, Simon RM, Friedewald WT, *et al*. Randomized clinical trials: perspectives on some recent ideas. *New Engl J Med* 1976, **295**, 74–80.
18. Sylvester R. Randomized clinical trials versus historical control studies. *Chem Oncol* 1980, **4**, 245–249.
19. Zelen M. A new design for randomized clinical trials. *New Engl J Med* 1979, **300**, 1242–1245.
20. Altman DG, Whitehead J, Parmar MKB, Stenning SP, Fayers PM, Machin D. Randomised consent designs in cancer clinical trials. *Eur J Cancer* 1995, **31A**, 1934–1944.
21. Pocock SJ, Lagakos SW. Practical experience of randomization in cancer trials: an international survey. *Br J Cancer* 1982, **46**, 368–375.
22. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975, **31**, 103–115.
23. Freedman LS, White SJ. On the use of Pocock and Simon's method for balancing treatment numbers over prognostic factors in the controlled clinical trial. *Biometrics* 1976, **32**, 691–694.
24. Aaronson NK, Ahmedzai S, Bergman B, *et al*. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993, **85**, 365–376.
25. Neymark N. *Assessing the Economic Value Of Anticancer Therapies. Recent Results in Cancer Research,* vol. 148. Berlin, Springer-Verlag, 1998.
26. De Gruttola VG, Clax P, DeMets DL, *et al*. Considerations in the evaluation of surrogate endpoints in clinical trials: summary of a National Institutes of Health Workshop. *Controlled Clin Trials* 2001, **22**, 485–502.
27. Parmar MKB, Machin D. *Survival Analysis: A Practical Approach*. Chichester, John Wiley and Sons, 1995.
28. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958, **53**, 457–481.
29. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 1966, **50**, 163–170.
30. Cox DR. Regression models and life tables. *J Roy Stat Soc B* 1972, **34**, 187–220.
31. Begg C, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomized controlled trials: the CONSORT Statement. *JAMA* 1996, **276**, 637–639.
32. Moher D, Schultz KF, Altman DG, for the CONSORT Group. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001, **357**, 1191–1194.
33. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994, **69**, 979–985.
34. Vaida F, Xu R. Proportional hazards model with random effects. *Stat Med* 2000, **19**, 3309–3324.
35. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991, **266**, 93–98.
36. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med* 1999, **18**, 695–706.
37. Fleming T, DeMets D. Monitoring of clinical trials: issues and recommendations. *Controlled Clin Trials* 1993, **14**, 183–197.
38. Smith MA, Ungerleider RS, Korn EL, Rubinstein L, Simon R. Role of independent data-monitoring committees in randomized clinical trials sponsored by the National Cancer Institute. *J Clin Oncol* 1997, **15**, 2736–2743.
39. Albert PS. Longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 1999, **18**, 1707–1732.
40. Curran D, Bacchi M, Hsu Schmitz SF, Molenberghs G, Sylvester RJ. Identifying the types of missingness in quality of life data from clinical trials. *Stat Med* 1998, **17**, 739–756.
41. Sylvester R, Collette L, Duchateau L. The role of meta-analyses in assessing cancer treatments. *Eur J Cancer* 2000, **36**, 1351–1358.
42. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. *J Clin Oncol* 2000, **18**, 2327–2340.
43. Rowinsky EK. The pursuit of optimal outcomes in cancer therapy in a new age of rationally designed target-based anticancer agents. *Drugs* 2000, **60**, 1–14.