ELSEVIER

European Journal of Cancer 41 (2005) 1261–1266

European Journal of Cancer

www.ejconline.com

Biological surrogate end-points in cancer trials: Potential uses, benefits and pitfalls

Rachel Cooper [a,*], Johannes H.A.M. Kaanders [b]

[a] Department of Clinical Oncology, Leeds Cancer Centre, Cookridge Hospital, Hospital Lane, Leeds LS16 6QB, UK
[b] Department of Radiation Oncology, Radboud University Medical Centre, Nijmegen, The Netherlands

Received 25 February 2005; accepted 2 March 2005
Available online 29 April 2005

Abstract

New technologies have led to the development of an increasing number of targeted therapies and interest in combining these with conventional therapy to provide individualised patient treatments. New drug or treatment regimens must, however, undergo rigorous testing under strictly controlled conditions before they can be adopted as standard. This can be expensive, time-consuming and inefficient. Surrogate end-points have been proposed as an alternative, which could be measured earlier or more conveniently than true end-points. The aim of this paper is to review the definition, advantages, disadvantages and potential pitfalls of biological surrogate end-points in the context of cancer treatment.

Keywords: Surrogate end-points; Biomarkers; Cancer

1. Introduction

With the development of new technologies including genome sequencing, DNA microarrays, proteomics and imaging modalities such as positron-emission tomography (PET), we have a greater understanding of tumour biology and behaviour. This, in turn, has led to the development of an increasing number of targeted therapies and interest in combining these with conventional therapy to provide individualised patient treatments. However, any new drug or treatment regimen must undergo rigorous testing under strictly controlled conditions before it can be adopted as standard. Testing usually takes the form of a series of phase I, II and III trials, each with well-defined end-points. In most phase III trials, efficacy over the standard treatment is proven by showing a statistical improvement in an outcome, such as survival or local control, or a reduction in toxicity, a process which is both time-consuming, expensive and in some cases inefficient. In order to overcome these problems, it has been proposed that surrogate end-points, which could be measured earlier or more conveniently, might be an alternative to true end-points [1]. In addition, there has been great interest in developing and incorporating biomarkers into clinical trials to aid in the selection of compounds for testing and defining appropriate patient groups for trials or treatment [2]. The incorporation of the measurement of biomarkers in prospective trials might be helpful in determining the mechanism of treatment effect, lack of effect or toxicity. With these points in mind, a Surrogate End-point Group was established as part of the European Organisation for Research and Treatment of Cancer (EORTC) Radiotherapy Translational Research Group. The aim of this group is to identify where surrogate end-points might be appropriately investigated or incorporated into the trials of the Radiotherapy Group.

* Corresponding author. Tel.: +44 0 113 3924346; fax: +44 0 113 3924361.
E-mail address: rachel.cooper@leedsth.nhs.uk (R. Cooper).

0959-8049/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2005.03.002

The aim of this paper is to review the definition, advantages, disadvantages and potential pitfalls of biological surrogate end-points.

2. What is the definition of a surrogate end-point?

Surrogate literally means 'to substitute for' [3]. Therefore, in the simplest terms a surrogate end-point is a measurement that can be substituted for a true end-point to predict either benefit (e.g., survival) or harm (e.g., late toxicity). However, the lack of consistency in defining surrogate end-points has led to confusion. Recently, the Biomarkers Definitions Working Group proposed a general definition of a surrogate end-point as a 'biomarker that is intended to substitute for a clinical end-point and is expected to predict clinical benefit (or harm or lack of benefit) based on epidemiological, therapeutic, pathophysiological, or other scientific evidence' [4]. Furthermore, any changes induced in the surrogate end-point by a treatment must accurately reflect changes in the true end-point [5].

There should be a clear distinction between surrogate end-point and surrogate marker. Bentzen *et al*. [6] distinguished between a surrogate marker and a surrogate end-point in the context of late effects of radiotherapy. They defined a surrogate marker as a biological effect of treatment that, if it occurs, changes the probability of an individual developing a late effect, whereas a surrogate end-point does not necessarily predict development of an effect at the individual level, but is an indicator of the toxicity of a treatment at the trial level. This definition of surrogate marker corresponds more closely with the definition of a biomarker as a characteristic that is measured and evaluated as an indicator of normal processes, pathogenic processes or as a response to treatment [4]. In fact, it has been proposed that the term 'surrogate marker' be avoided as this suggests that the substitute is for a 'marker', rather than for a clinical end-point [4]. Clearly, this is an area which requires further clarification.

In 1989, Prentice [7] published an important paper which set down strict statistical criteria to define surrogate end-points. He defined a surrogate end-point as a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true end-point. That is, if we reject the null hypothesis that the surrogate end-point is associated with the treatment, meaning there is an association then there is most likely to be an association with the true end-point. Furthermore, a surrogate end-point must fully 'capture' the relationship between the treatment and the true end-point. This definition with respect to evaluation of surrogate end-points is discussed further below.

Various biological and clinical phenomena could potentially serve as surrogate end-points. These include molecular markers (specific mutations in cancer-related genes, gene expression products), cellular and nuclear phenomena (proliferation, apoptosis, DNA ploidy), serum markers (prostate-specific antigen (PSA), carcinoembryonic antigen (CAE), CA125, β human chorionic gonadotropin (βHCG), α fetoprotein (αFP), tumour characteristics detected by (functional) imaging (magnetic resonance imaging (MRI), PET) and clinical assessments (tumour response, time to progression).

3. What is the definition of a true end-point?

A true end-point is any characteristic or variable that reflects how a patient functions, feels or survives [8]. In radiotherapy trials, this usually means survival, local recurrence or development of toxicity. Time is usually measured from the start of treatment or date of pathological confirmation of disease. End-points such as time to progression and response (complete, stable or progression) are not true end-points but ought to be considered as surrogate end-points. Although, whether in most circumstances they qualify as such given the definitions and criteria discussed in this paper is highly disputable.

4. When can a biological surrogate end-point be used as a substitute for a true end-point?

Prior to evaluation of a potential surrogate end-point statistically, three criteria need to be satisfied: (i) is the potential surrogate associated with the true end-point biologically; (ii) is the treatment associated with the potential surrogate end-point; and (iii) does the potential surrogate mediate the effect of the treatment on the true end-point [8]?

To satisfy the first criteria we need to show that there is good biological evidence or a sound rationale to suppose that the surrogate is associated with the true end-point. Data to support this are most likely to be obtained from pre-clinical and animal studies as well as previous retrospective and epidemiological studies.

To satisfy the second criteria, we must be able to show that there is some relationship between the surrogate end-point and the treatment, that is the treatment changes the surrogate end-point. This information might be obtained from previous studies or relatively smaller studies designed to answer this question.

Finally, we need to show whether the effect of treatment on the true end-point is mediated via the surrogate end-point. This is important in situations where the effect of the treatment on the true end-point is mediated through mechanisms other than the surrogate or where the effect is mediated through a number of mechanisms

including the surrogate [9]. In these situations, the surrogate may not 'capture' the whole beneficial effect of the treatment or, perhaps more importantly, only capture the beneficial effects but not the toxic effects. Situations in which the disease pathogenesis and mechanisms of action of the treatment are well understood greatly enhance the likelihood of finding useful surrogate end-points.

Serum PSA is already widely used as a surrogate end-point, both for the early detection of prostate cancer and for evaluation of treatment outcome. There is debate, however, about the validity of this end-point and the exact nature of its relationship with the true end-points: occurrence of clinically significant prostate cancer and cancer-specific mortality, respectively [10]. Due to the protracted nature of prostate cancer and life expectancies of many years even for patients with overt distant metastases, there is a strong desire to shorten the time required to assess the value of new treatments.

It can be readily agreed that the PSA-assay satisfies the two first criteria. Serum PSA-levels are associated with the load of PSA-producing tumour cells, which to a large degree determines the prostate cancer-specific survival probability. Clinical studies have confirmed this relationship [11]. Reduction of the tumour burden after treatment will be reflected in a reduction in the PSA-level. The effect of treatment on survival is, however, not mediated by PSA production. This reduces the likelihood that the treatment effect is fully captured by the PSA assay. This is illustrated by the fact that not all prostate cancers and, maybe more importantly, not all sub-populations of cells of the same cancer express PSA equally. In addition, it has been demonstrated that the association between PSA failure and survival is not the same for all treatments [10].

It is widely believed that tumour cells treated with anticancer agents die by apoptosis and that tumours that do not readily undergo apoptosis are resistant to treatment. Under this assumption, apoptosis as a mechanism of action of the treatment is likely to satisfy the third criteria and would be of potential interest as a surrogate end-point. However, although apoptosis is causally related with a true end-point such as local tumour control, other mechanisms may play a role as well. This was demonstrated in a study of human cervix tumour cell lines [12]. Nine tumour cell lines were analysed which demonstrated different levels of background and radiation-induced apoptosis. Interestingly, there was a significant inverse correlation between the level of radiation-induced apoptosis and necrosis, indicating that cell kill occurs both through apoptosis and through necrosis and that either of the two can be the dominant mechanism of action in different tumours. This example illustrates that, although the true end-point is mediated via the surrogate end-point, the surrogate does not completely capture the effect of the treatment.

5. Evaluation of surrogate end-points

Any proposed surrogate end-point that fulfils the above criteria must be formally evaluated before it can replace a true end-point in either the clinical or trial situation [13]. This has been an area of much debate since the publication of Prentices' criteria in 1989. Several modifications of the original criteria have been proposed [13–17], however there is still no consensus [18].

In the criteria set out by Prentice (see above), a perfect surrogate would be one where all of the effect of treatment on the true end-point is mediated through the surrogate end-point. However, it is more likely that the treatment works via a number of different mechanisms not all of which are mediated through the surrogate [9,18]. Furthermore Prentice's criteria have been criticised as being both stringent and difficult to verify [5]. Therefore, a number of subsequent publications proposed refinements of the original Prentice criteria. One approach was to use the 'proportion explained' (PE) which describes the proportion of the treatment effect mediated by the surrogate [16]. In this model, a good surrogate end-point would be one where PE approached unity suggesting that most of the effect of the treatment is mediated through the surrogate. Subsequently, two further quantities were suggested, the 'relative effect' (RE) and the 'adjusted association' (AA) [13]. The relative effect relates to the ratio of the overall treatment effect on the true end-point to the effect on the surrogate end-point. If at the population level the RE is one then it is said to be a perfect surrogate. The adjusted association refers to the association between both end-points at the individual level after adjusting for the effect of treatment [13]. Using these parameters, a quantitative assessment of the value of the surrogate can be provided in addition to an estimate of the effect of a treatment on the true end-point given the effect on the surrogate end-point [13]. For a surrogate to be valid at the trial level, data from a number of different trials using the same class of treatment and measuring the same surrogate end-point must be analysed. If the squared correlation (R^2) of the effect of treatment on the surrogate plotted against the effect of treatment on the true end-point for each trial is high then the surrogate is said to be valid at the trial level [14,15]. A criticism of this approach, however, is that there is loss of precision using a surrogate end-point compared with the true end-point and uncertainty surrounding comparisons of different treatment regimens that are thought to work by the same mechanism [17]. An example is the meta-analysis by Buyse *et al.* [19] on the relation between tumour response (surrogate end-point) to first-line chemotherapy and survival (true end-point) in advanced colorectal cancer. Their analysis confirmed that an increase in tumour response rate translates into an increase in overall survival. However, there was significant loss of

precision. A treatment that lowered the odds of failure to respond by 50% would be expected to decrease the odds of death by only 6%. The R^2 of the regression line between the effects of treatment on response and the effects of treatment on survival was 0.38 (CI: 0.09–0.68) (Fig. 1). This means that less than half of the variability in the treatment effects on survival could be explained by treatment effects on response. In practice, this implies that knowledge that a treatment has benefits on tumour response does not allow very accurate prediction of the ultimate benefit on survival at the individual trial level.

If a surrogate end-point is to be used to compare treatments, for example in a randomised trial, in each of the randomisation arms a high correlation is needed between the effect of the treatments on the surrogate end-point and on the true end-point. However, simple correlation within each of the randomisation arms does not make a valid surrogate end-point. Fig. 2 is a reproduction from a publication by Baker and Kramer [20] graphically illustrating this pitfall. The graph applies to a hypothetical randomised trial and shows a high correlation between true and surrogate outcomes for a control group and an experimental group. Despite good correlations, the slopes of the lines for the two groups differ. As shown, the mean surrogate outcome in the control group is higher than the surrogate outcome in the experimental group. However, because of the difference in slopes, this is the reverse for the true outcome, demonstrating that in this situation use of the surrogate end-point would lead to opposite (and wrong) conclusions. A typical example of this phenomenon was demonstrated in the Radiation Therapy Oncology Group (RTOG)-study 92-02, which compared short *versus* long-term adjuvant androgen deprivation in locally advanced prostate cancer [21]. This study showed that the relative risk of prostate cancer related death after

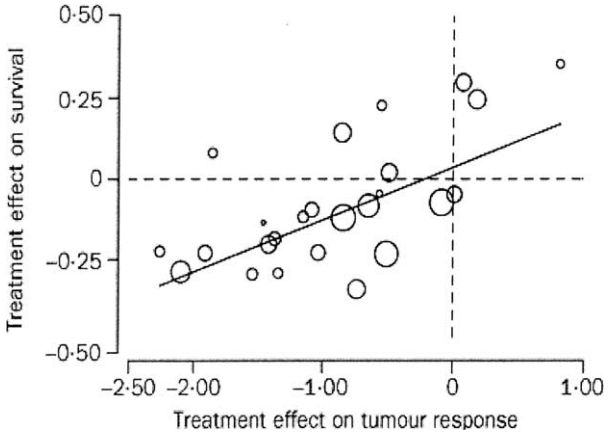

Fig. 1. Treatment effects on survival (log hazard ratio) *versus* treatment effects on tumour response (log odds ratio). Each circle represents a randomised trial of first-line chemotherapy in advanced colorectal cancer, the area of which is proportional to the number of observations in the trial. $R^2 = 0.38$. Reproduced with permission [19].

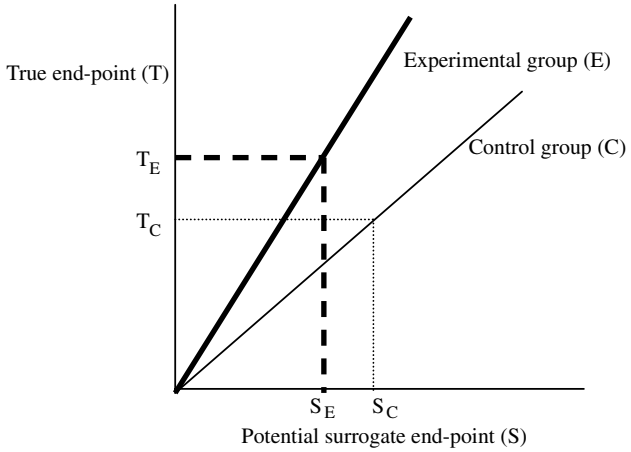

Fig. 2. Results of a hypothetical randomised clinical trial of an experimental (E) *versus* control (C) treatment. Although there is a good correlation between the true end-point (T) and surrogate end-point (S) for both treatments the mean true outcome for the experimental group is higher than that of the control group. However, the opposite is true for the surrogate end-point. In this situation, the use of the surrogate end-point would lead to the opposite (and wrong) conclusions. Reproduced with permission [20].

PSA failure was 1.5 times greater after long-term androgen deprivation compared with short-term treatment. Thus, the validity of PSA as a surrogate end-point is weakened when it is used to compare the effects of different durations of androgen ablation.

In order to estimate RE with good precision, a large number of observations are needed and multiple studies required in order to differentiate between individual and trial level associations [13]. Unfortunately, these are both situations we are trying to avoid by introducing the use of a surrogate end-point.

Finally, a surrogate end-point should have a high degree of accuracy reflected by both a high positive and negative predictive value [6,9].

6. What trials are needed to validate a biological surrogate end-point?

To validate potential surrogate end-points they should be integrated into studies with well-defined true clinical events as the primary end-point. Within these studies, standards must be set for assessment of the candidate surrogate. These include appropriate timing of the surrogate assay, quality control and standards for interpretation of the results. Also, the size of these studies must be large enough and follow-up must be sufficiently long to evaluate the true end-point and to assemble sufficient data to assess performance of the surrogate. To show a consistent performance, information across a variety of studies is required. Furthermore, the validity of a surrogate must be established for every intervention. As previously shown, the association between the

surrogate end-point and the true end-point may not be the same for all treatments. An approach to this problem is to consider studies of a 'class' of biologically comparable interventions [8]. If the surrogate end-point has been validated for a class of interventions, we can be reasonably confident that a new member of that class will show very similar associations between surrogate and true end-points. To estimate the potential value of a surrogate for a certain new intervention in an early stage, it should be integrated already in the very early pre-clinical and clinical testing of the new therapeutic strategy.

The irony here is that the large and extended studies required to validate a surrogate end-point are precisely the type of laborious studies that it was hoped that surrogates would replace.

7. What are the potential advantages of surrogate end-points?

There are several potential advantages of using surrogate end-points in clinical trials. Clinical trials, especially phase III trials, are often long and expensive. A surrogate end-point could potentially allow a trial to be completed earlier and with a reduced number of subjects and, therefore, at less cost. Surrogate end-points may also be attractive in phase II trials to help provide an estimate of the effect of the new treatment under investigation. This might allow introduction of beneficial treatments at an earlier stage or rejection of treatments with no benefit before embarking on costly and time-consuming phase III trials [18]. In addition, use of surrogate end-points in early clinical trials may assist to better select patient- or tumour categories that are most likely to benefit from new treatments. A proper selection of the study population can reduce the required sample size in a randomised trial. Surrogate end-points may also be more conveniently or more frequently measured [1]. Surrogates may also be used in a situation where we have competing risks or the true end-point is complicated by the use of other treatments, a common situation in trials of cancer treatment [7]. In certain studies, surrogate end-points or biomarkers might be useful in establishing the mechanism of action of a treatment or provide better understanding of the underlying mechanism of the disease process [18]. Finally, surrogate end-points may be used in a broader context, not only for assessment of the benefits of a treatment but also as indicators of late toxicity. Proper evaluation of late morbidity after radiotherapy requires extended follow-up because of the long latent period and very large patient numbers because of the low prevalence of severe late effects. These are exactly the conditions where surrogate end-points can be of benefit.

An ideal surrogate would also have a long lead-time, which might allow modification of ongoing treatment or initiation of alternative treatments [6].

8. What are the potential disadvantages of biological surrogate end-points?

Unfortunately, despite the recent interest and extensive data on potential surrogate end-points currently being generated from studies on the molecular biology and genetics of disease processes and the effect of treatment, very few surrogate end-points have been introduced into routine clinical practice. This reflects the difficulties associated with using surrogate end-points. One of the most important problems relates to the complexity of the place of the surrogate end-point in the relationship between the treatment and the true end-point. In the ideal situation all of the effect of the treatment on the true end-point would be mediated through the surrogate. However, this is probably never the situation [9]. First, the surrogate might not be on the pathway between the treatment and the true end-point or it may be on the pathway but insensitive to the effects of treatment. Second, the treatment effect might be mediated by a number of pathways only one of which includes the surrogate end-point, which additionally may play only a minor role in the treatment effect. The treatment might only affect the disease process mediated by the surrogate marker although there might be other processes affecting the true end-point independent of the treatment. Finally, the treatment might have other effects independent of the disease process. In any of these situations, the surrogate end-point is likely to either underestimate (or in the worse case completely miss) the effect of the treatment or overestimate any benefit.

As discussed above, a surrogate end-point should totally capture any effect of the treatment. However, it is unlikely that any one end-point would capture both the beneficial and harmful effects of treatment. Disease-related surrogates will generally underestimate toxic effects and it may well be that a surrogate does not detect potential side-effects of a treatment which would render a new treatment unacceptable [22].

Furthermore, as different treatments are likely to work through different mechanisms and the effect of a treatment is likely to be different in different malignancies, the applicability of a surrogate end-points from one situation to another must be considered carefully. Finally, it is likely that a panel of surrogate end-points will be required to capture fully the complex nature of any disease–treatment interaction.

9. Conclusion

The inherent difficulties and expense of large-scale phase III trials with end-points such as survival or late toxicity has led to increasing interest in performing trials with surrogate end-points that potentially offer a quicker and less expensive solution. However, despite

the ever-increasing information on potential surrogate end-points in cancer, very few are in routine clinical use. This might reflect a number of problems including the lack of appropriate trials incorporating surrogate end-points and the uncertainty around the reliability of surrogate end-points compared with true end-points. This has been compounded by the failure of a number of surrogate end-points in non-oncological trials [9,22].

A surrogate end-point is likely to capture only one aspect of a treatment effect, which could either under- or over-estimate potential benefit whilst missing harmful effects. Due to the complicated nature of cancer–treatment interactions, it is therefore likely that a panel of surrogate end-points (possibly in the form of multiple biomarkers) will be required to predict the full effect of the treatment. Furthermore, in order to validate such markers, large and usually long studies will be required and their applicability to different treatment regimens or cancers will always be an area of uncertainty.

However, despite these problems, research should continue into surrogate end-points. In the phase II setting they might be useful to determine the best candidate treatment to go onto large phase III trials [18]. Furthermore, in the context of phase III trials, information from surrogate end-points might be useful in 'strengthening' the analysis of the true end-point, the concept of auxiliary end-points [5] or elucidating the mechanism of a treatment [9].

Conflict of interest statement

None declared.

References

1. Ellenberg SS, Hamilton JM. Surrogate endpoints in clinical trials: cancer. *Stat Med* 1989, **8**, 405–413.
2. Rolan P. The contribution of clinical pharmacology surrogates and models to drug development: a critical appraisal. *Brit J Clin Pharmacol* 1997, **44**, 219–225.
3. Karol Sikora. Surrogate endpoints in cancer drug development. *Drug Discov Today* 2002, **7**, 951–956.
4. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Br. J Clin Pharmcol Ther* 2001, **69**, 89–95.
5. Fleming TR, Prentice RL, Pepe MS, *et al*. Surrogate and auxiliary endpoints in clinical trials, with potential applications in cancer and AIDS research. *Stat Med* 1994, **13**, 955–968.
6. Bentzen SM, Dörr W, Mitchell S, *et al*. Normal tissue effects: reporting and analysis. *Semin Radiat Oncol* 2003, **13**, 189–202.
7. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989, **8**, 421–440.
8. Schatzkin A, Gail M. The promise and peril of surrogate end points in cancer research. *Nat Rev Cancer* 2002, **2**, 19–27.
9. Fleming TR, DeMets DL. Surrogate end points in clinical trials. Are we being misled? *Ann Intern Med* 1996, **125**, 605–613.
10. Sandler HM, DeSilvio ML. Surrogate end points for prostate cancer: what is prostate-specific antigen telling us. *J Natl Cancer Inst* 2003, **95**, 1352–1353.
11. D'Amico AV, Moul JW, Caroll PR, *et al*. Surrogate end point for prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *J Natl Cancer Inst* 2003, **95**, 1376–1383.
12. Sheridan MT, West CM. Ability to undergo apoptosis does not correlate with the intrinsic radiosensitivity (SF2) of human cervix tumor cell lines. *Int J Radiat Oncol Biol Phys* 2001, **50**, 503–509.
13. Buyse M, Molenberghs G. criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 1998, **54**, 1014–1029.
14. Buyse M, Molenberghs G, Buzykowski T, *et al*. The validation of surrogate endpoints in meta-analyses of randomised experiments. *Biostatistics* 2000, **1**, 49–67.
15. Daniels M, Hughs MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med* 1997, **16**, 1965–1982.
16. Freedman LS, Graubard BI, Schatzkin A. Satistical validation of intermediate endpoints for chronic disease. *Stat Med* 1992, **11**, 167–178.
17. Gail MH, Pfeiffer R, Van Houwelingen H, *et al*. On met-analytic assessment of surrogate outcomes. *Biostatistics* 2000, **1**, 231–246.
18. 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. *Control Clin Trials* 2001, **22**, 485–502.
19. Buyse M, Thirion P, Carlson RW, *et al*. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. *Lancet* 2000, **356**, 373–378.
20. Baker SG, Kramer BS. A perfect correlate does not make a surrogate. *BMC Med Res Methodol* 2003, **3**, 16.
21. Sandler HM, Pajak TF, Hanks GE, *et al*. Can biochemical failure (ASTRO definition) be used as a surrogate endpoint for prostate cancer survival in phase III localized prostate cancer clinical trials? Analysis of RTOG protocol 92-02. *Proc Am Soc Clin Oncol* 2003, **22**, 381.
22. Echt DS, Liebson PR, Mitchell LB, *et al*. Mortality and morbidity in patients receiving encainide, flecainide or placebo. The cardiac arrhythmia suppression trial. *N Engl J Med* 1991, **324**, 781–788.