# *The Roles of Short Term Endpoints in*

# *Clinical Trial Planning and Design*

## Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

http://people.bath.ac.uk/mascj

*VIB, Adaptive Trials 2009 Conference*

*Brussels, December 2009*

# Examples of short term endpoints

Short term responses can often provide an indication of how a treatment will ultimately affect the primary, long term endpoint.

|  | *Short term endpoint* | *Long term endpoint* |
|---|---|---|
| *Altzheimer's disease* | Change in mental acuity score after 2 months | Change in mental acuity score after 6 months |
| *Osteoporosis* | Bone mineral density | Fracture |
| *Cancer* | Progression free survival | Overall survival |
| *HIV infection* | CP4 count or viral load | Progression to AIDS |

A short term endpoint can guide the direction of a trial or of a sequence of trials.

More radically, a short term endpoint may be used as a substitute or "surrogate" for the real endpoint.

# Benefits of using short term endpoints

A short term endpoint may be more convenient to use if it is:

*More rapidly observed,*

*Less expensive to measure.*

A short term endpoint may provide good insight into a treatment's effect if it:

*Is more immediately affected by the treatment,*

*Has a high incidence rate, while the long term endpoint concerns a rare event.*

***But, remember that the goal is to demonstrate a treatment effect on the long term endpoint.***

# Potential uses for short term endpoints

In this talk, I shall consider a range of uses of short term responses.

In some cases, a short term endpoint is simply used in place of the long term endpoint:

**1.** A surrogate endpoint may replace the long term endpoint — if it is accepted that demonstrating an effect on the surrogate endpoint is sufficient evidence for treatment approval.

**2.** A short term endpoint can be used in a Phase II study for establishing proof of concept or dose-finding.

# Potential uses of short term endpoints

In other cases, a rapidly observed response is used in interim decision making, but the final decision is based on the long term response:

**3.** In an adaptive Phase II design, the rule for dose allocation may use short term responses for subjects who do not yet have the long term response.

**4.** In a group sequential test for a delayed response, a short term response can be incorporated in interim analyses to gain information from subjects who have been treated but not yet produced a response.

**5.** In an adaptive Phase II / III design, a decision on the population to be investigated in Phase III may be based on a short term response observed in Phase II, then the treatment effect on this population is evaluated in Phase III using the long term endpoint.

# 1. Using a surrogate endpoint to prove a treatment effect

Reference: Chapter 2 of *The Evaluation of Surrogate Endpoints*, Burzykowski, Mohlenbergs and Buyse (20005).

Regulatory agencies allow surrogate endpoints in certain circumstances.

*USA*

An accelerated approach based on surrogate endpoints may be allowed for indications with no effective therapy.

Approval is then conditional: the drug can be marketed while further studies assess long term benefit. In 1996, the FDA stated

> "FDA believes that for many cancer therapies it is appropriate to utilize objective evidence of tumor shrinkage as a basis for approval, allowing additional evidence of increased survival and/or improved quality of life . . . to be demonstrated later."

# Successful use of surrogate endpoints in cancer

The achievement of a complete remission substantially improves prognosis in haematological malignancies.

In solid tumours, the relationship between response and survival is less clear.

Some examples where response rate has been used for accelerated approval in the USA are:

| *Treatment* | *Indication* |
|---|---|
| Docetaxel | Second-line metastatic breast cancer |
| Irinotecan | Second-line metastatic colorectal cancer |
| Capecitabine | Refractory metastatic breast cancer |

# Using a surrogate endpoint to prove a treatment effect

*Europe*

The European Union also has an "accelerated approval" mechanism.

Legislation allows marketing authorisation, provided the applicant agrees to undertake studies that will be the basis for a later review of treatment benefits.

Thus, surrogate endpoints can be used in oncology as long as there is confirmation of an effect on the "true endpoint" in the post-authorisation phase.

Shorter term endpoints in Phase III oncology trials include:

*Progression-free survival,*

*Response rate,*

*Symptom control/quality of life.*

# Using a surrogate endpoint to prove a treatment effect

*Japan*

Objective response rate has played a central role in oncology drug approval and cytotoxic drugs can be approved based on tumour shrinkage in Phase II studies.

Initial approval is conditional on re-examination of safety and efficacy of the drug 4 to 10 years after marketing (with at least two trials with a survival endpoint).

*ICH Guidelines*

"In practice the strength of evidence for surrogacy depends on

    (i) The biological plausibility of the relationship,

    (ii) Demonstration in epidemiological studies of the prognostic value of the surrogate endpoint,

    (iii) Evidence from clinical trials that treatment effects on the surrogate correspond to effects on the clinical outcome."

# Using a surrogate endpoint to prove a treatment effect

ICH condition (iii) requires:

> Evidence from clinical trials that treatment effects on the surrogate
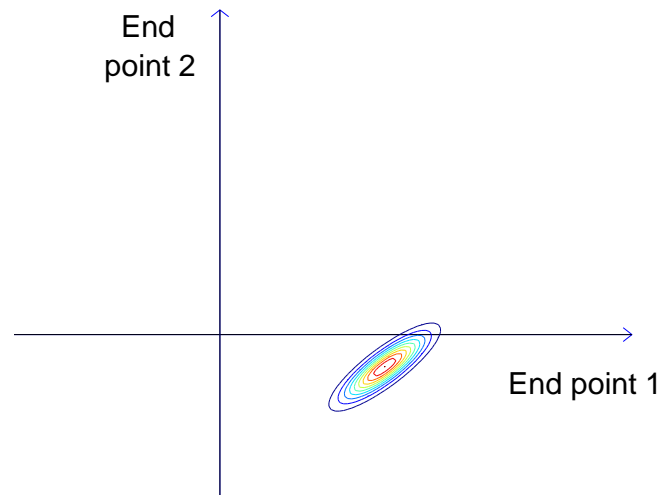> correspond to effects on the clinical outcome.
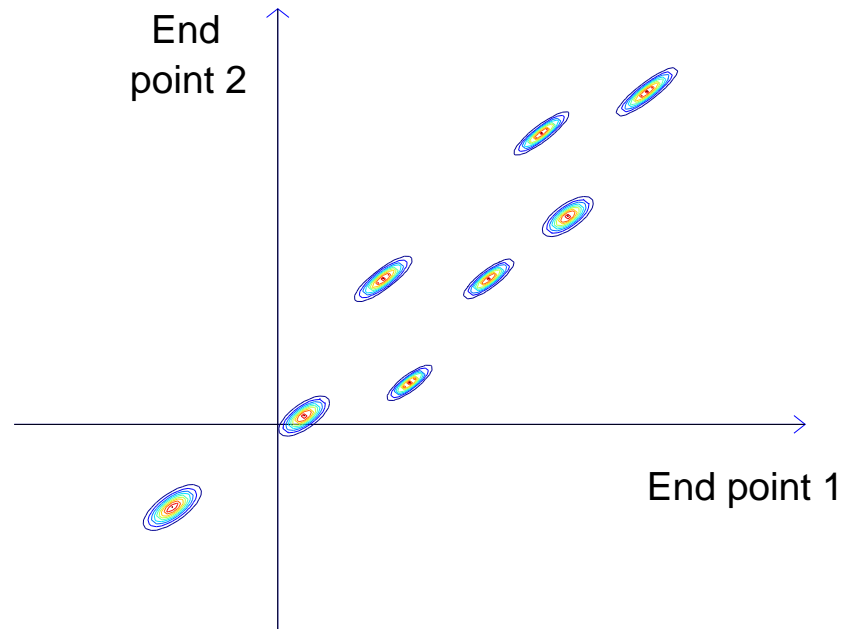


A sizable set of trials for similar treatments on the same condition is needed!

# Using a surrogate endpoint to prove a treatment effect

It is important to note that

*"A correlate is not necessarily a surrogate."*

One could have, for example,



For more formal criteria for "surrogate endpoint evaluation," see Prentice (*Statistics in Medicine*, 1989).
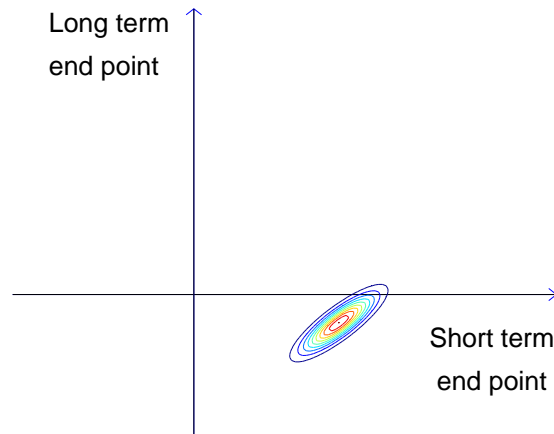
# 2. Using a short term endpoint in a Phase II trial

## (1) Proof of concept

Investigators seek evidence in a Phase IIa trial that, at a particular dose level, the treatment has an effect on the short term endpoint.

If so, they will proceed to dose finding in Phase IIb.

The reasoning here assumes that a treatment effect on the short term endpoint implies an effect on the longer term, primary endpoint.

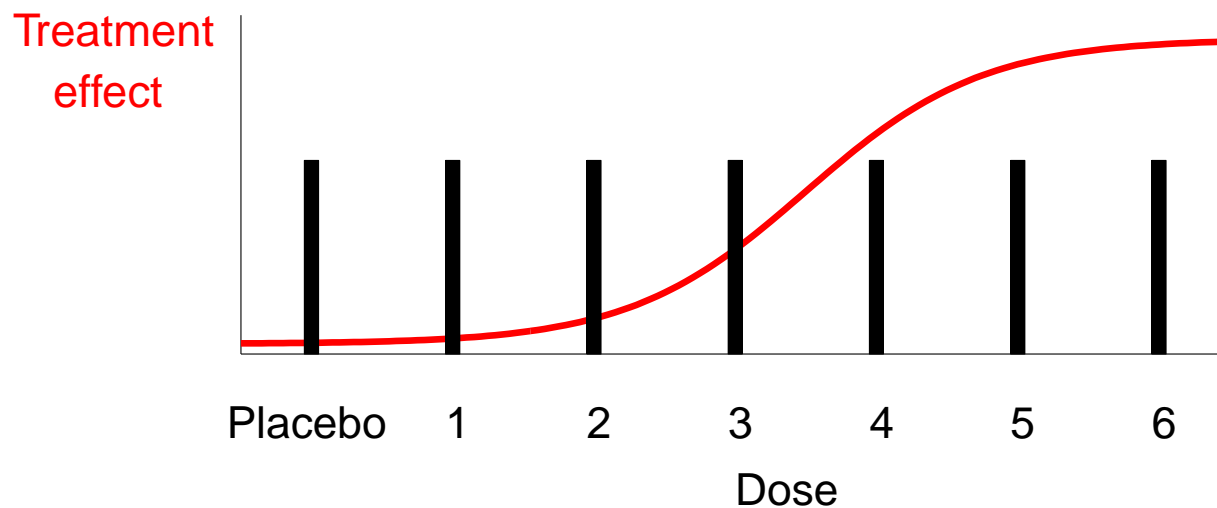So, the preceding example is not envisaged:

# Using a short term endpoint in a Phase II trial

## (2) *Dose finding*

A parallel arm trial compares patient responses over a range of dose levels in order to select the most suitable dose.

The objective is to find a dose which is both effective and safe.
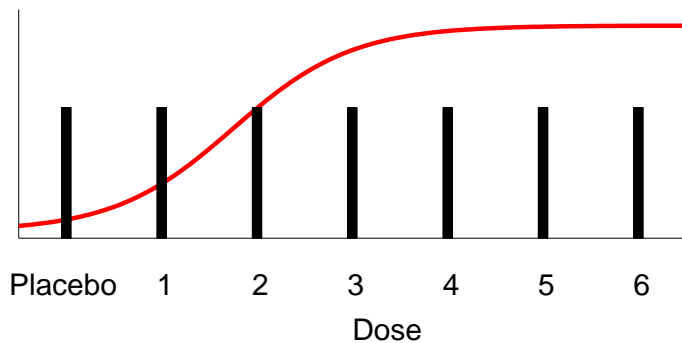


When we use a short term response for dose finding, we implicitly assume the dose-response curve for the long term endpoint will have similar characteristics.
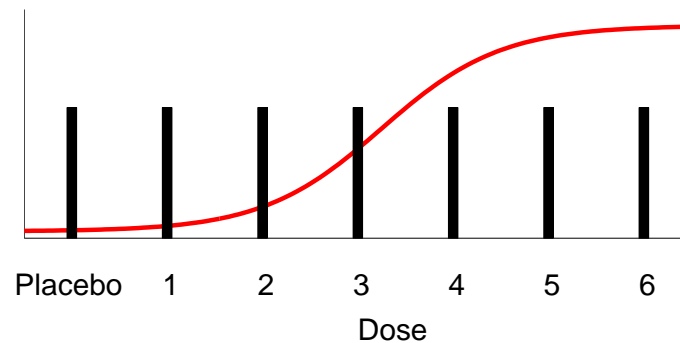
# Using a short term endpoint in a Phase II trial

It would be unfortunate if dose response were to follow different patterns for short term and long term endpoints:

**Short-term endpoint**

**Long-term endpoint**



Here, results for the short term endpoint could lead to selection of dose 2 or 3, whereas observing the long term endpoint would lead to choosing dose 4 or 5.

It is standard to use Phase II data to plan a Phase III trial. Investigators should be aware of the underlying assumptions about the relationship between endpoints.

# Using a short term endpoint in a Phase II trial

A PhRMA Adaptive Progams Workstream has been set up, chaired by Carl-Fredrik Burman, to investigate the joint design of Phase II and Phase III trials.

The result from Phase II could be:

A go/no-go decision (proof of concept),

A choice of dose or doses to study in Phase III.

In considering the efficiency of an overall "design strategy", one is led to consider how well results from Phase II predict the likelihood of success in Phase III.

If Phase II uses a different endpoint from Phase III, a key issue is the quality of the Phase II endpoint as a "surrogate" for the real Phase III endpoint.

But, how often are sequences of Phase II and Phase III trials planned with clear knowledge about the relationship between such endpoints?

## 3. Combined use of short term and long term endpoints in a Phase II dose finding trial

Adaptive dose finding designs allocate patients to dose levels in a way that provides information about the most interesting part of the dose response curve.
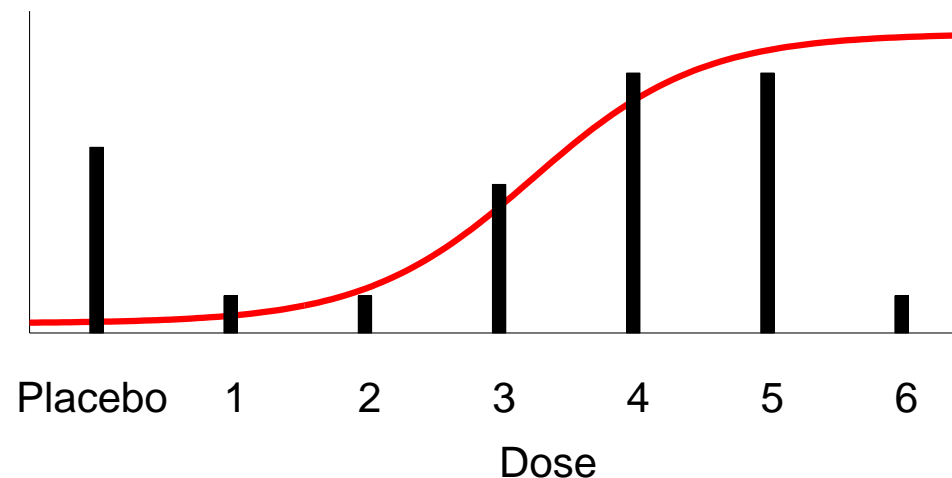


The ASTIN trial (Krams et al., *Stroke*, 2003) used *Bayesian adaptive allocation.*

The endpoint here was a **long term** one: change from baseline in Scandinavian Stroke Scale score after 90 days.

# Use of short term endpoints in the ASTIN trial

The optimal allocation of patients to doses depends on the true dose response curve — which can only be estimated from current data at interim stages.
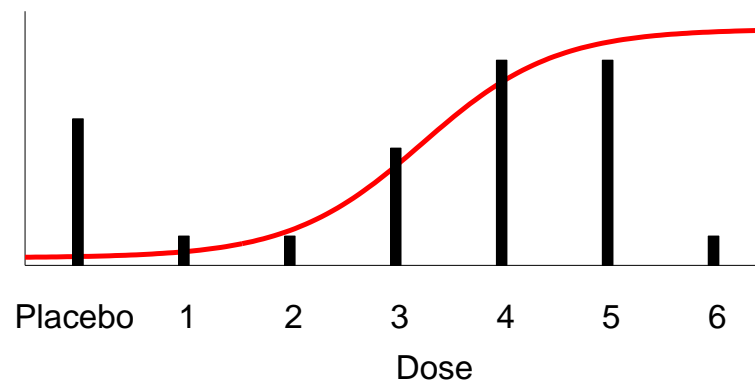


At an interim point, some patients have not been in the trial for 90 days but they do have, say, 30 day or 60 day scores.

Fitting a joint model to short term and long term responses allows patients with only short term responses to contribute to estimation of the dose response curve for the 90 day endpoint.

# 4. Combined use of short term and long term endpoints in a group sequential test

Consider a group sequential trial testing $H_0: \theta \leq 0$ against $\theta > 0$, where $\theta$ is the treatment effect for an experimental therapy compared to placebo.
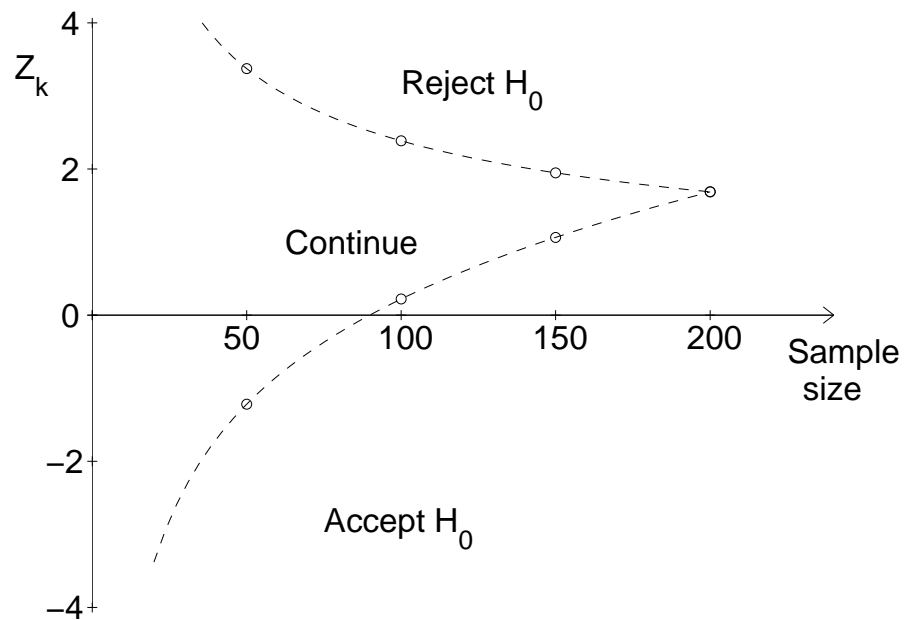


A group sequential design allows the trial to stop early when appropriate, with proper control of type I and II error rates.

## Using short and long term endpoints in a group sequential test

Suppose the primary endpoint is a long term response.

At an interim analysis of a group sequential test, there will be "pipeline" subjects who have been treated but not yet responded.

In recent work with Lisa Hampson, we have formulated group sequential designs which deal properly with a delayed response.
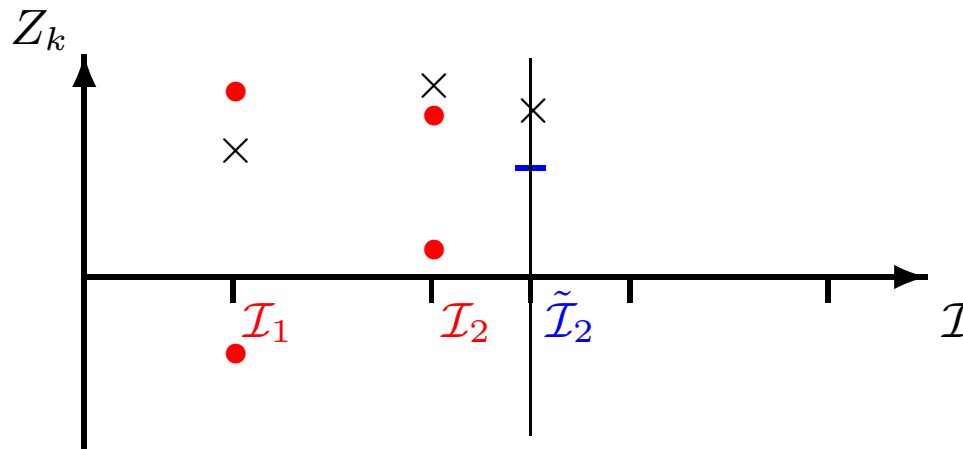


Recruitment can cease at the $k$th analysis with information $\mathcal{I}_k$ for $\theta$, but the final decision is made with extra "pipeline" data bringing the information up to $\tilde{\mathcal{I}}_k$.

# Using short and long term endpoints in a group sequential test

Even with optimised designs, little value is obtained from treated subjects who have not responded when recruitment terminates. This erodes the benefits of lower average sample size than a fixed sample size design.

*Solution:* Fitting a joint model to short term and long term responses allows pipeline patients to contribute to estimation of the long term endpoint.

If the within patient correlation between short term and long term responses is $\rho$, a pipeline patient with a short term response contributes $\rho^2$ of the information of a fully observed patient.

So $\rho = 0.7$ gives 50% value to these subjects and $\rho = 0.9$ gives 80% value.

NB The final inference is about the *long term response*.

We do not need to assume treatment benefit on the short term endpoint translates into a positive effect on the long term endpoint.

*It is fine for the short term endpoint to be a correlate, not a surrogate.*

# 5. A Phase II / III design with population selection

Suppose the full patient population  F  has a subgroup  S  in which a treatment effect is thought to be particularly likely.

In a seamless Phase II / III trial, Phase II data can be used to define the population in which a treatment effect will be sought in Phase III.

Three options are possible:

      F      Full population,

      S      Sub-population,

      FS    Co-primary populations, F and S.

***An oncology example***

Jenkins, Stone and Jennison (2009) describe a trial design where the choice of Phase III population is based on Progression Free Survival (PFS) data in Phase II.

Overall survival data are obtained for both Phase II and Phase III subjects.

# A Phase II / III oncology trial design with population selection

The decision about which population(s) to consider in Phase III is based on estimated hazard ratios for PFS in Phase II.

|  |  | Estimated hazard ratio for PFS in F | |
|---|---|---|---|
|  |  | $< 0.8$ | $> 0.8$ |
| Est. hazard ratio for PFS in S | $< 0.6$ | Co-primary analysis, test $H_0^F$ and $H_0^S$ | Subgroup only, test $H_0^S$ |
| Est. hazard ratio for PFS in S | $> 0.6$ | Full population only, test $H_0^F$ | Stop for futility |

Null hypotheses $H_0^F$ and $H_0^S$ state treatment has no effect on overall survival in the full population and subgroup, respectively.

F = full population, S = subgroup.

# A Phase II / III oncology trial design with population selection

The selected hypotheses ($H_0^F$, $H_0^S$ or both) are tested by combining p-values from logrank tests based on Phase II and Phase III data.

A multiple comparisons method (e.g., a closed testing procedure, Marcus et al, *Biometrika*, 1976) is needed to adjust for the data-dependent choice of hypotheses.

For this two-stage adaptive design, a Bauer & Köhne combination test (*Biometrics*, 1994) can be used to combine p-values from the two phases.

*Care is necessary in defining p-values so that they are not subject to bias due to the adaptive choice of hypotheses.*

| | | |
|---|---|---|
| *Phase II patients* | Overall survival (during Phase II) | Overall survival (during Phase III) |
| *Phase III patients* | | Overall survival (during Phase III) |

# A Phase II / III design with population selection

*A non-biased definition of p-values*

|  | Overall survival (during Phase II) | Overall survival (during Phase III) |
|---|---|---|
| *Phase II patients* | Overall survival (during Phase II) | Overall survival (during Phase III) |
| *Phase III patients* |  | Overall survival (during Phase III) |

Calculate the "stage 1" p-value from the logrank test for *all* the overall survival data of Phase II patients, observed over both Phase II and Phase III.

Calculate the "stage 2" p-value from the logrank test for overall survival data of Phase III patients observed in Phase III.

Using Phase III follow-up of Phase II subjects in a "stage 2" p-value can cause bias.

*For this design to be of value, a treatment effect on PSF should imply a very likely effect on overall survival.*

# Conclusions

1.  Under certain conditions, a short term endpoint can be used as a surrogate for the real, long term endpoint.

Broadly speaking, it must be known that a treatment effect on the short term endpoint implies an effect on the real endpoint.

2.  A similar relationship should be present between Phase II and Phase III endpoints if the results of Phase II are to be of value in deciding whether or not to proceed to Phase III, or in choosing the treatment dose to study in Phase III.

3.  Less stringent conditions apply to the use of a short term endpoint in certain adaptive or group sequential designs.

If the short term endpoint is to serve as a precursor to a long term endpoint, it is sufficient to have high within patient correlation between these endpoints.