Issues in Combining Phase IIb and Phase III Clinical Trials

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

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

Workshop on Adaptive Designs, International Biometric Society

Berlin, September 2006

Seamless transition between Phase IIb and Phase III

1. Separate trials for Phase IIb (dose finding) and Phase III (confirmation)

   Making use of a break between phases

   Implications of delay in reaching a positive conclusion

2. Seamless transition: Joint planning of Phase IIb and Phase III trials

   Writing a single protocol

   “Non-statistical” gains of a combined approach

   Statistical methodology for combining phases

   Efficiency gains from using Phase IIb data in a combined analysis

3. Further good practices for Phase IIb/Phase III trials

1. Running separate Phase IIb and Phase III trials

Phase IIb

The trial compares several dose levels of a treatment with a control in order to select a dose and provide evidence of improvement against the control.

A Phase III

The trial is run as a confirmatory study to demonstrate superiority against control of the treatment selected in Phase IIb.

Stages:

Write Phase IIb protocol, seek ethical & regulatory approval, (FDA, IRBs, . . .)

Run Phase IIb, analyse data, reach conclusions.

Write Phase III protocol, seek ethical & regulatory approval, (FDA, IRBs, . . .)

Run Phase III, analyse data, reach final conclusion.

Planning the Phase III trial

Planning the Phase III trial after Phase IIb allows investigators to make use of information gained in Phase IIb.

They may decide to modify:

Treatment definition,

Target population,

Primary endpoint,

Sample size.

Positive results in Phase IIb will help recruitment for participation in Phase III.

But, planning and gaining approval for the Phase III trial can be time-consuming.

If the final outcome is positive, the sooner this conclusion is reached, the better.

2. Joint planning of Phase IIb and Phase III trials

Requirements

A single protocol for the combined Phase IIb and Phase III trials.

Rules for a committee managing the trials to follow as they:

Decide whether to proceed to Phase III,

Select the treatment, with regard to both efficacy and safety outcomes,

Respond to information from Phase IIb, e.g., using estimated response variance to set Phase III sample size.

NB, expect everyone else to be blinded to the Phase IIb results.

Potential benefits

Eliminating the “white space” between phases,

Gaining efficiency from using Phase IIb data in the Phase III analysis.

When to combine Phase IIb and Phase III data?

No:

1. A Phase IIb trial providing evidence of efficacy of the selected treatment relative to the control can sometimes be used in place of a second confirmatory study.

If this is a possibility, the benefit will be much greater than that gained from using some of the Phase IIb data in Phase III.

2. A Phase IIb trial may use a more rapid patient response, such as a bio-marker or an earlier measurement of the long term outcome.

If the primary endpoint for Phase III is not available for Phase IIb patients, the data cannot easily be combined.

Possibly yes when:

Treatment, patient population and endpoint are the same in both cases.

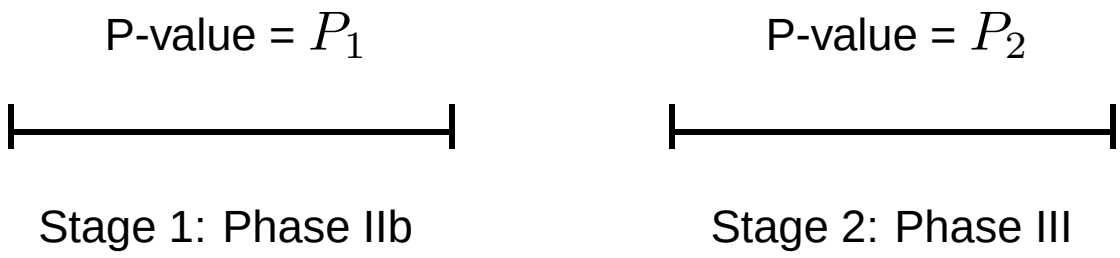
The efficiency gains are sufficient to make the exercise worthwhile.

3. Combining Phase IIb and Phase III data

Case 1: Comparing n dose levels* vs *control in Phase IIb

Data from Phase IIb and Phase III trials can be analysed together using a combination test, just as in flexible adaptive designs.

For example, following Bauer and Köhne (*Biometrics*, 1994), one calculates a P-value from both "stages" and combines these values.



Bauer and Köhne combine P_1 and P_2 through Fisher's test, using the fact that

$$-\log(P_1\,P_2) \;\sim\; \tfrac{1}{2}\,\chi^2_4$$

if P_1 and P_2 have $U(0,1)$ distributions.

Combining Phase IIb (n doses *vs* control) and Phase III

Bauer and Köhne, method (a)

Let θ_i, $i = 1, \ldots, n$, denote the effect size of dose level i *vs* the control treatment.

Obtain P_1 from a test of H_{01}: $\theta_1 = \ldots = \theta_n = 0$ using Phase IIb data.
One might, for example, test for a positive trend in effect as dose increases.

Select dose level i^*, then use Phase III data to test H_{02}: $\theta_{i^*} = 0$, yielding P_2.

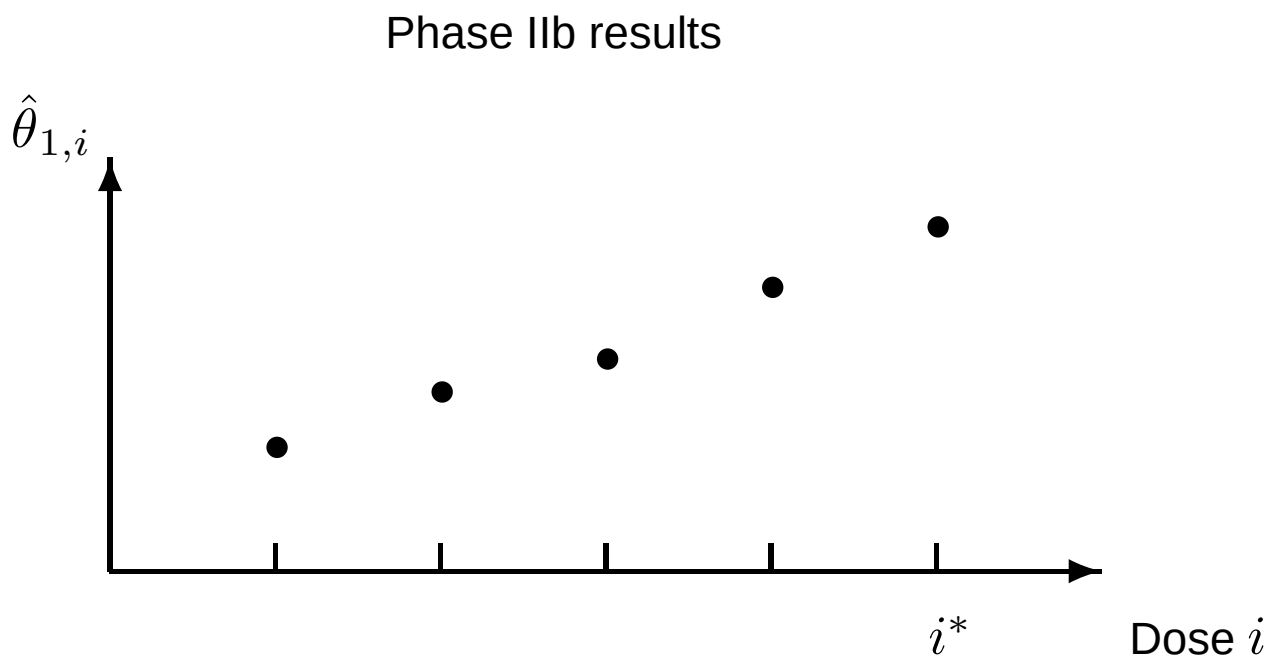
NB, P_2 is still $U(0, 1)$ under H_{02}, even if the Phase III design is adapted to Phase IIb findings.

The combined test based on $\log(P_1\, P_2)$ has overall null hypothesis

$$H_0 \;=\; H_{01} \cap H_{02} \;=\; H_{01}.$$

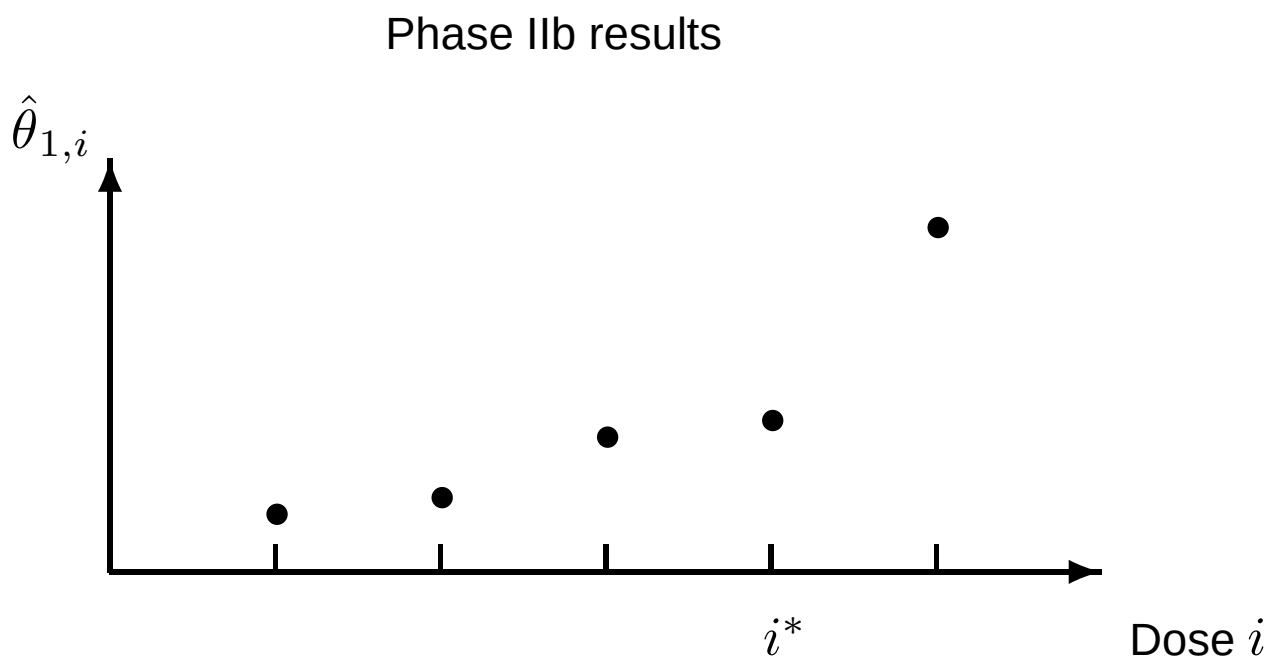
Does rejection of H_0 imply $\theta_{i^*} > 0$?

Bauer and Köhne, method (a)



Clear evidence of a simple dose response relationship in Phase IIb supports rejection of H_{02}: $\theta_{i^*} = 0$, as well as of H_{01}: $\theta_1 = \ldots = \theta_n = 0$.

Bauer and Köhne, method (a)



Evidence of a dose response relationship is dominated by results at the highest dose, but this dose is not selected for Phase III due to safety problems.

Rejecting $H_0 = H_{01} \cap H_{02}$ does *not necessarily* imply $\theta_{i^*} > 0$.

Combining Phase IIb (n doses *vs* control) and Phase III

Bauer and Köhne, method (b)

See also Bretz, Schmidli et al. (*Biometrical Journal*, 2006).

Procedure

Select dose level i^* to advance to Phase III.

Hope to reject H_{i^*}: $\theta_{i^*} = 0$ at significance level α, with allowance for the multiple comparisons involved in Phase IIb.

Formally

Define H_i: $\theta_i = 0$ for $i = 1, \ldots, n$.

We want a procedure controlling the **family wise error rate**, i.e., for all $\{\theta_i\}$,

$$Pr\{\text{Reject any true } H_i\} \leq \alpha.$$

Then, the probability of falsely claiming significance for the selected i^* is at most α.

Bauer and Köhne, method (b)

To protect the family wise error rate, use the ***Closure Principle:***

For each subset I of $\{1, \ldots, n\}$, define a level α test of the intersection hypothesis

$$H_I = \cap_{i \in I} H_i.$$

The simple hypothesis H_j: $\theta_j = 0$ is rejected if, and only if, H_I is rejected for every set I containing index j.

Proof of overall error rate

Let $\tilde{I}$ be the set of all true hypotheses H_i.

For a "family wise" error to be committed, $H_{\tilde{I}}$ must be rejected. Since $H_{\tilde{I}}$ is true,

$$Pr\{\text{Reject } H_{\tilde{I}}\} = \alpha$$

and so the probability of a family wise error is no greater than α.

Bauer and Köhne, method (b)

Using the Closure Principle in combination Phase IIb/Phase III trials

Phase IIb

Observe estimated treatment effects $\hat{\theta}_{1,i}$, $i = 1, \ldots, n$.

Select treatment i^* to go forward to Phase III.

Treatment i^* will have a high estimate $\hat{\theta}_{1,i^*}$ and good safety outcomes.

Phase III

Test treatment i^* against control.

In order to reject H_{i^*}: $\theta_{i^*} = 0$, we need to reject each intersection hypothesis H_I with $i^* \in I$ at level α.

Here, $H_I = \cap_{i \in I} H_i$ states that $\theta_i = 0$ for all $i \in I$.

Bauer and Köhne, method (b)

Testing an intersection hypothesis H_I***:*** $\theta_i = 0$ ***for all*** $i \in I$

a) Need to test an intersection hypothesis.

b) Need to combine data from two stages, Phase IIb and Phase III.

Take problem (b) first

We can use a combination test, following Bauer and Köhne (1994).

Denote the P-value for testing H_I in Phase IIb by $P_{1,I}$.

Denote the P-value for testing H_I in Phase III by $P_{2,I}$.

Reject H_I if

$$-\log(P_{1,I}\, P_{2,I}) > \frac{1}{2}\, \chi^2_{4,\,1-\alpha}.$$

Bauer and Köhne, method (b)

Testing an intersection hypothesis

(a) Testing H_I is most complex in Phase IIb

Suppose we calculate a P-value, $P_{1,i}$, for each H_i: $\theta_i = 0$.

Using the Bonferroni inequality, the overall P-value for testing H_I is m times the minimum $P_{1,i}$ over $i \in I$, where m is the number of indices in I.

Schmidli, Bretz et al. (*Biometrical Journal*, 2006) propose using Simes' (*Biometrika*, 1986) modification of the Bonferroni inequality:

Let $P_{1,(k)}$, $k = 1, \ldots, m$, denote the m P-values in increasing order.

Then the P-value for testing H_I is

$$P_{1,I} = \min_{k=1,\ldots,m} (m\, P_{1,(k)}/k).$$

Bauer and Köhne, method (b)

Testing an intersection hypothesis

Using Simes' method:

The P-value for testing H_I, where I has m elements, is

$$P_{1,I} = \min_{k=1,\ldots,m} (m\, P_{1,(k)}/k).$$

If treatment i^* has the highest $\hat{\theta}_{1,i}$ and smallest P-value of all n treatments, we have $P_{1,(1)} = P_{1,i^*}$ in any set I containing i^*.

The term $m\, P_{1,(k)}/k$ with $k = 1$ becomes $m\, P_{1,i^*}$, the usual "Bonferroni adjusted" version of P_{1,i^*}.

Simes' method allows other low P-values to reduce the overall result: if a second treatment performs well, $P_{1,(2)}/2$ may be smaller than P_{1,i^*}, reducing $P_{1,I}$.

Bauer and Köhne, method (b)

Testing an intersection hypothesis

(a) Testing H_I in Phase III

In order to reject H_{i^*}: $\theta_{i^*} = 0$, we need to reject each H_I with $i^* \in I$.

Only treatment i^* is studied in Phase III, so a test of such an H_I using Phase III data is based on $\hat{\theta}_{1,i^*}$ — and there is just one such test.

Hence, all H_I of interest have a common P-value in Phase III, $P_{2,I} = P_{2,i^*}$.

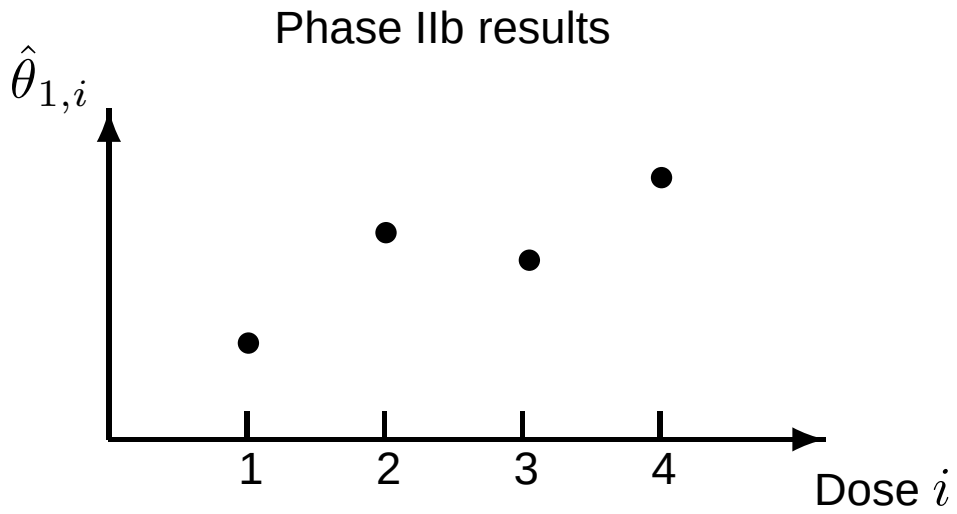
Since the combination test of H_I is based on $\log(P_{1,I}\, P_{2,I})$, rejection of H_{i^*} depends on the highest value of $\log(P_{1,I}\, P_{2,i^*})$.

This comes from the H_I with highest $P_{1,I}$, so the key statistic from Phase IIb is:

$$\max_I P_{1,I} \text{ over sets } I \text{ containing } i^*.$$

Bauer and Köhne, method (b)

Simes' test: Example



Select $i^* = 4$ for Phase III.

Suppose $P_{1,1} = 0.2,\ P_{1,2} = 0.04,\ P_{1,3} = 0.05,\ P_{1,4} = 0.03.$

For sets I containing $i^* = 4$, maximum $P_{1,I}$ comes from $I = \{1, 3, 4\}$.

Ordered P-values are $P_{1,(1)} = 0.03,\ P_{1,(2)} = 0.05,\ P_{1,(3)} = 0.2$:

$$P_{1,I} \ = \ \min_{k=1,\ldots,3} (3\, P_{1,(k)}/k) \ = \ 3 \times 0.05/2 \ = \ 0.075.$$

Combining Phase IIb (n doses *vs* control) and Phase III

Summary of Bauer and Köhne, method (b)

In Phase IIb:

Select treatment i^* and carry forward $\max_{I:i^* \in I} P_{1,I}$.

In Phase III:

Test treatment i^* against control and find P_{2,i^*}.

Overall:

Combine the two P-values to see if H_{i^*}: $\theta_{i^*} = 0$ is rejected.

Flexibility:

Treatment i^* can be selected for efficacy, safety, or other factors

— not necessarily the treatment with maximum $\hat{\theta}_{1,i}$.

Efficiency:

Phase IIb data increases power or reduces Phase III sample size.

Combining Phase IIb (n doses *vs* control) and Phase III

Alternative methods

Two-stage procedures for treatment selection and testing include proposals by:

Thall, Simon and Ellenberg (*Biometrika*, 1988),

Schaid, Wieand and Therneau (*Biometrika*, 1990),

Stallard and Todd (*Statistics in Medicine*, 2003).

These tests are not presented in terms of the closure principle, but they can be interpreted in that framework.

We shall focus first on the approach of Thall, Simon and Ellenberg (TSE).

See Jennison and Turnbull's (*Biometrical Journal*, 2006), discussion of the Bretz, Schmidli et al. papers for further details.

Thall, Simon and Ellenberg

Phase IIb

Take m_1 observations per treatment and control.

Denote estimated effect of treatment i against control by $\hat{\theta}_{1,i}$ and let the maximum of these be $\hat{\theta}_{1,i^*}$.

If $\hat{\theta}_{1,i^*} < C_1,$ stop and accept H_0: $\theta_1 = \ldots = \theta_k = 0$,

if $\hat{\theta}_{1,i^*} \geq C_1,$ select treatment i^* and proceed to Phase III.

Phase III

Take m_2 observations on treatment i^* and the control.

Combine data in $T_{i^*} = (m_1\,\hat{\theta}_{1,i^*} + m_2\,\hat{\theta}_{2,i^*})/(m_1 + m_2)$.

If $T_{i^*} < C_2,$ accept H_0,

if $T_{i^*} \geq C_2,$ reject H_0 and conclude $\theta_{i^*} > 0$.

Thall, Simon and Ellenberg

Type I error and power requirements imply values for m_1, m_2, C_1 and C_2.

Type I error

Treatment i^* is said to be "chosen" if

Treatment i^* is selected at the end of Phase IIb, and

H_0 is rejected in favour of $\theta_{i^*} > 0$ in the final analysis.

The type I error rate is

$$Pr\{\text{Any experimental treatment is "chosen"}\}$$

under H_0: $\theta_1 = \ldots = \theta_k = 0$.

Thall, Simon and Ellenberg

Power

Power depends on the full vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$.

Any treatment with $\theta_i \geq \delta_2$ is said to be "acceptable".

Consider cases of $\boldsymbol{\theta}$ where:

At least one treatment is acceptable,

No θ_i lies in the interval $(\delta_1,\ \delta_2)$.

The power function is $Pr_{\boldsymbol{\theta}}\{\text{An acceptable choice is made}\}$.

Thall, Simon and Ellenberg

Power

TSE show power is minimized in cases described above under the *least favourable configuration:*

$$\theta_1 = \ldots = \theta_{k-1} = \delta_1 \quad \text{and} \quad \theta_k = \delta_2.$$

They call this configuration $\boldsymbol{\theta}^*$ and set their power condition at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

Numerical integration under H_0 and $\boldsymbol{\theta}^*$ enables parameters m_1, m_2, C_1 and C_2 to be found satisfying type I error and power conditions.

Tests minimizing expected sample size averaged over these two cases are found by searching feasible parameter combinations.

Thall, Simon and Ellenberg

The TSE method has family wise error rate α

TSE set their type I error probability α under the global null hypothesis H_0: $\theta_1 = \ldots = \theta_k = 0$.

Family wise error rate concerns the probability of "choosing" a treatment with $\theta_i \leq 0$ under any vector $\boldsymbol{\theta}$, including cases where $\theta_j > 0$ for some indices j.

Starting from the case $\theta_1 = \ldots = \theta_k = 0$, one can show that moving some effects to values $\theta_i < 0$ decreases the family wise error probability.

Then, increasing other values to $\theta_j > 0$ reduces the error rate as there are fewer "unacceptable" treatments and the chance of selecting one of these is decreased.

We conclude that the TSE method protects the family wise error rate at level α.

Thall, Simon and Ellenberg

The TSE method follows the Closure Principle

Define a test of H_I for I containing i^* as follows:

If the process stops at Phase IIb, accept H_I.

When proceeding to Phase III define

$$T_I \;=\; T_{i^*} \;=\; \frac{m_1\,\hat{\theta}_{1,i^*} + m_2\,\hat{\theta}_{2,i^*}}{m_1 + m_2},$$

and reject H_I for $T_I > c_I$, accept H_I otherwise.

The critical value c_I is set to make this a level α test of H_I. The highest c_I is for $I = \{1, \ldots, n\} = I_n$, say.

Rejecting H_{I_n} implies rejecting every other H_I with $i^* \in I$.

Equating C_2 and c_{I_n}, we see that "choosing" i^* is equivalent to rejecting H_{i^*} according to the Closure Principle with these tests of intersection hypotheses H_I.

Thall, Simon and Ellenberg

Remarks on the TSE method

The single summary of Phase IIb data is $\hat{\theta}_{1,i^*}$. This is combined with Phase III data in the overall estimate of θ_{i^*} using the final sufficient statistic for θ_{i^*}.

The critical value for the final test statistic allows for selecting treatment i^* as the best performer in Phase IIb.

The null distribution taken for $\hat{\theta}_{1,i^*}$ is that of the maximum of n estimated effects when the true effects are all zero.

Adding flexibility:

The TSE method can be used, as defined, when a treatment i^* is selected with $\hat{\theta}_{1,i^*}$ less than the highest $\hat{\theta}_{1,i}$. The type I error rate is then met conservatively.

Alternatively, the above definitions of T_I, etc., can be extended to create a less conservative procedure via the Closure Principle.

Efficiency gains from combining Phase IIb and Phase III

Comparisons have been made of the total sample size in

Separate Phase IIb and Phase III trials *vs*

Combined design with Phase IIb data used at the end of Phase III.

Bretz et al. (2006) consider examples where sample size per treatment in Phase IIb is equal to sample size per treatment in Phase III.

The combined study saves 30% of the total sample size for selecting one of $n = 2$ treatments and testing *vs* control.

But, perhaps the Phase IIb trial could count as a supporting study instead?

Todd and Stallard (*Drug Information Journal*, 2005) present an example where sample size per treatment is 25 in Phase IIb and 1400 in Phase III.

Here, savings can be at most 2% of the total sample size!

Combining Phase IIb (n doses *vs* control) and Phase III

Comparison of methods: Bauer and Köhne (b), TSE, etc.

Further investigations are needed to compare efficiency of methods using

different summary statistics from Phase IIb,

different methods for combining Phase IIb and Phase III data.

Other variations on these methods are possible

More than one treatment may be carried forward to Phase III.

Sequential monitoring:

inferior treatments may be eliminated early in Phase IIb,

an early decision, positive or negative, may be reached in Phase III

— **with a greater efficiency gain than from combining Phase IIb/Phase III.**

Combining Phase IIb (n doses *vs* control) and Phase III

Selected references

Allowing more than one treatment to progress to Phase III:

Schaid, Wieand and Therneau (*Biometrika*, 1990).

Group sequential analysis of combined Phase IIb and Phase III data (with adjustment for treatment selection in Phase IIb):

Stallard and Todd (*Statistics in Medicine*, 2003).

Elimination procedures and adaptive treatment allocation:

Paulson (*Annals of Mathematical Statistics*, 1964),

Robbins and Siegmund (*JASA*, 1974),

Jennison, Johnstone and Turnbull (*Purdue Symposium*, 1982).

Combining Phase IIb (n doses *vs* control) and Phase III

Totally seamless procedure

In principle, one can combine

Interim monitoring,

Treatment elimination,

Carrying several treatments into Phase III,

Early stopping for a final decision.

Then, the distinction between Phases IIb and III is blurred and a totally seamless procedure emerges.

Complex calculations would be required to define such a procedure meeting specified error rates and to evaluate its properties.

It would help to identify the key components of an overall efficient procedure.

4. Combining Phase IIb and Phase III data

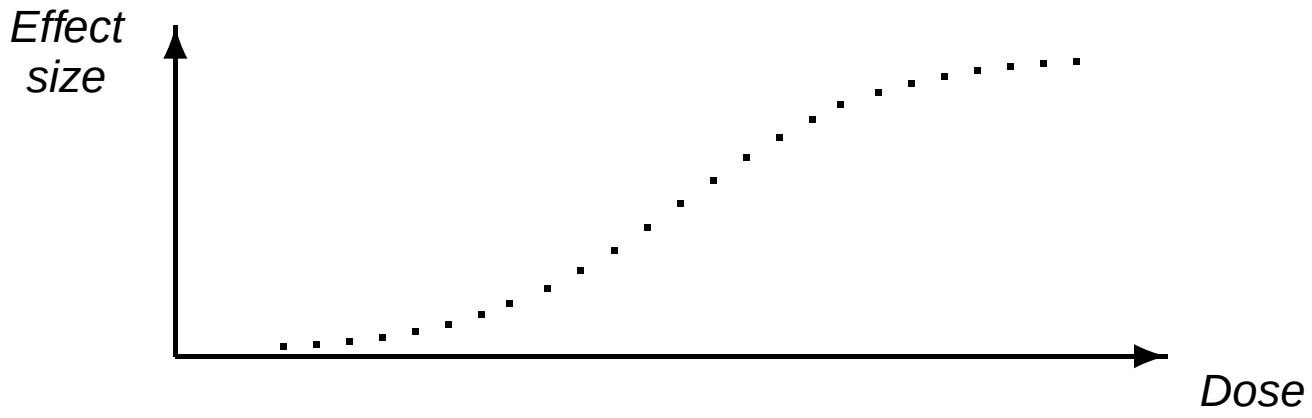
Case 2: Modelling dose response* vs *control in Phase IIb

Suppose it is allowable to assume a model for dose response in Phase IIb.

It is then much simpler to test

$$H_{i^*}: \text{ No treatment effect at the selected dose, } i^*$$

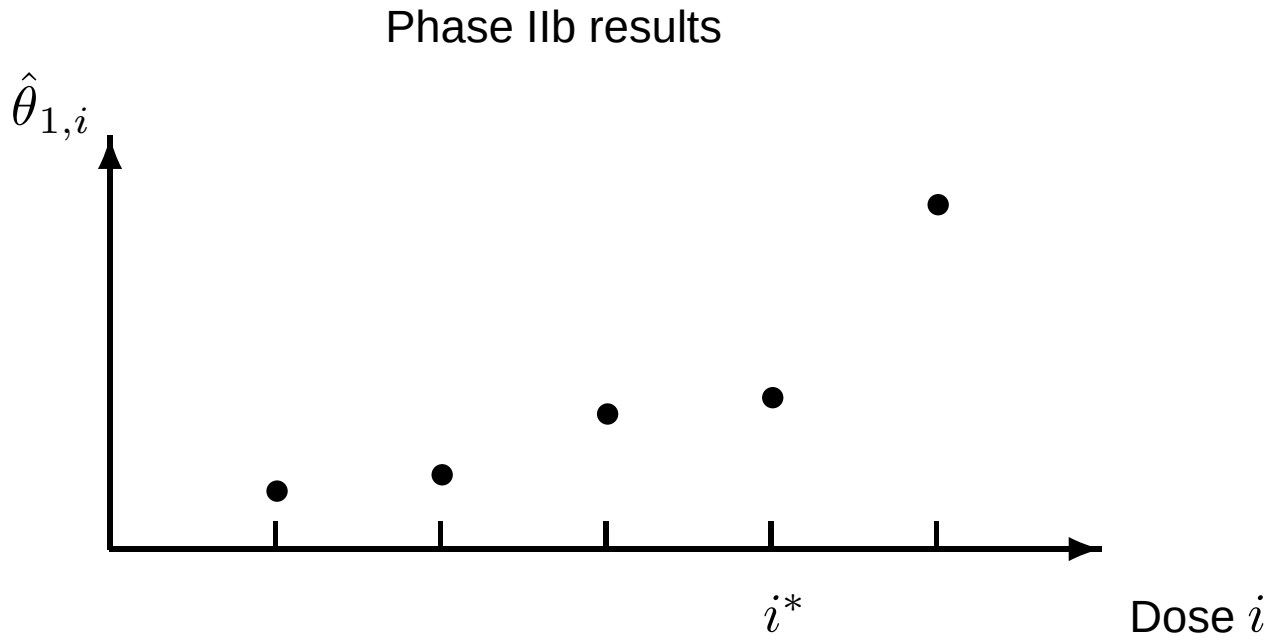
since typical models have a parameter θ in the dose response equation such that $\theta > 0$ implies H_{i^*} is is not true.



This seems to cut past the multiplicity question in Phase IIb too easily!

Modelling dose response in Phase IIb

Previous example:



The assumed model may mean that any positive trend implies positive treatment effects at all dose levels.

But in this example, the effect size at dose i^* could still be very slight.

Modelling dose response in Phase IIb

Research agenda

Using an appropriate dose-response suitable model should aid effective Phase IIb study design.

Integrating results from a model based Phase IIb trial into a combined analysis with Phase III data is desirable.

If only results at the selected dose i^* are permitted to contribute to a combined Phase IIb/Phase III analysis, special designs are called for.

The ASTIN trial (Krams et al. *Stroke*, 2003) is well worth studying for its use of modelling in the on-going design of a Phase IIb trial.

Simpler variants of such a design could make it more feasible to have a combined analysis of Phase IIb data with results of a subsequent Phase III trial.

Using a rapidly observable endpoint in Phase IIb

Research agenda

Different endpoints may be used in Phase IIb and Phase III.

A more rapidly available response (surrogate endpoints, bio-markers) may be used to select the treatment for which a long-term response is tested in Phase III.

How should efficient designs be derived in this case?

When is it useful to follow up (some) Phase IIb subjects to observe the long-term endpoint while the Phase III trial runs its course?

Does the delay involved in waiting for a long-term response change the nature of "efficient" trial designs?

The important questions in this area involve practical issues as much as statistical methodology.