Seamless Phase IIb / Phase III Clinical Trials

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

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

Föreningen för Medicinsk Statistik

Göteberg,

September 24, 2007

Plan of talk

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

3. Statistical methodology for combining phases

   Efficiency gains from using Phase IIb data in a combined analysis

4. 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 and regulatory approval, (FDA, IRBs, . . .)

Run Phase IIb, analyse data, reach conclusions.

Write Phase III protocol, seek ethical and 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:

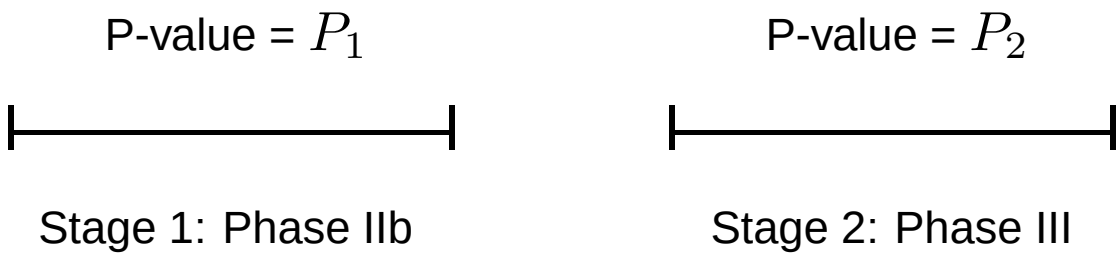
Treatment, patient population and endpoint are the same in both cases.

The efficiency gains are sufficient to make the exercise worthwhile.

3. Combining Phase IIb and Phase III data

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 independent $U(0,1)$ distributions.

Combining Phases IIb and III

Bauer and Köhne, method (a)

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

Obtain P_1 from a test of H_{01}: $\theta_1 \leq 0, \ldots, \theta_k \leq 0$ 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}: $\theta_{i^*} \leq 0$, yielding P_2.

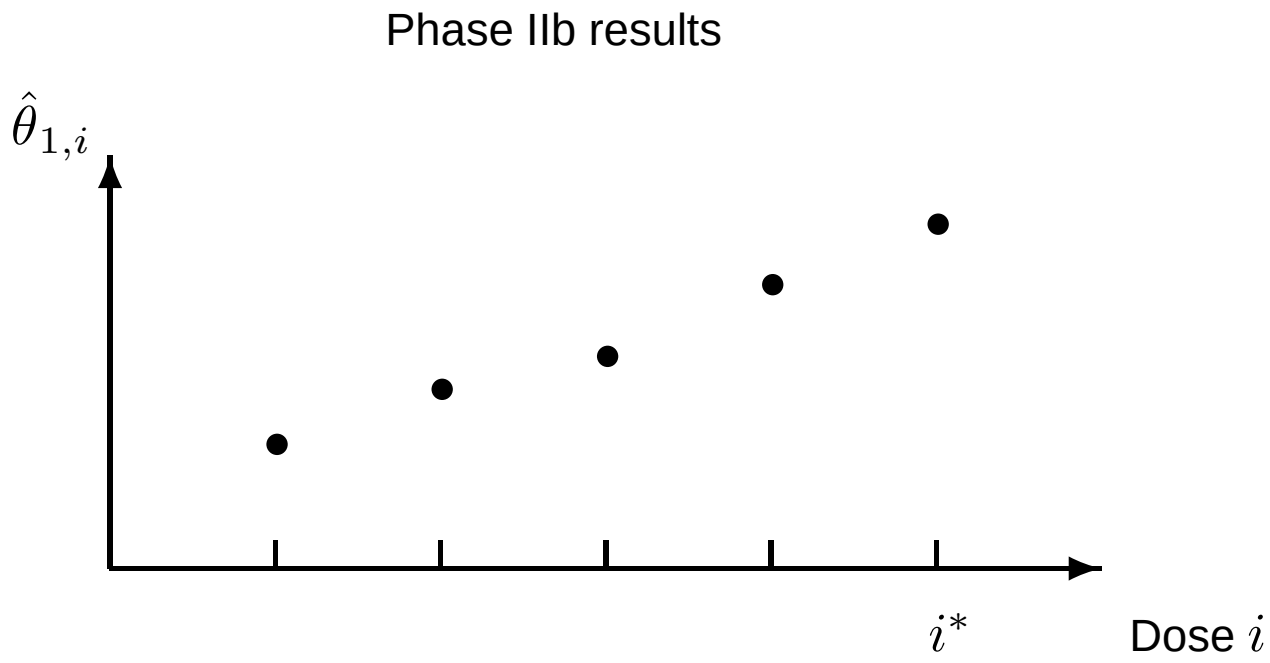
NB, P_2 is still $U(0,1)$ under $\theta_{i^*} = 0$, even if the Phase III design is adapted to Phase IIb findings.

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

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

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

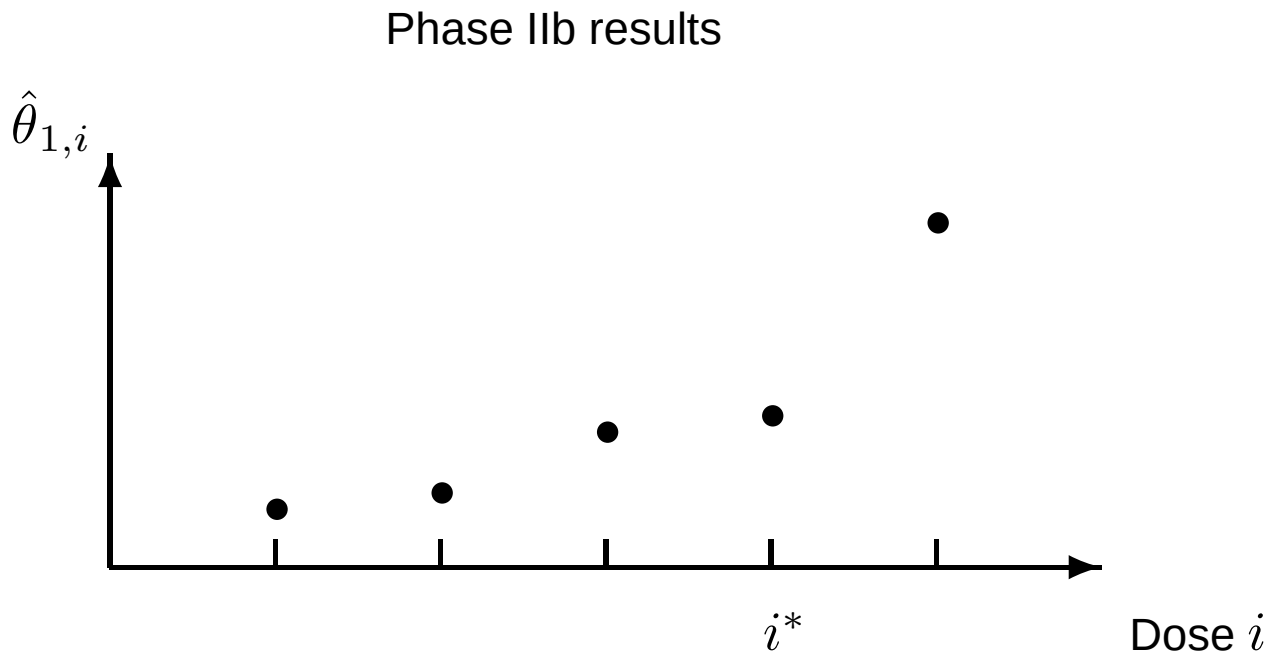
Combining tests of two null hypotheses



Evidence of a simple dose response relationship in Phase IIb appears to support rejection of H_{02}: $\theta_{i^*} \leq 0$, as well as of H_{01}: $\theta_1 \leq 0, \ldots, \theta_k \leq 0$.

But the picture is not always as clear as this . . .

Combining tests of two null hypotheses



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.

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

Combining Phases IIb and 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^*}: $\theta_{i^*} \leq 0$ at significance level α, with allowance for the multiple comparisons involved in Phase IIb.

Formally

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

We shall define a procedure controlling the **familywise error rate**.

Then, for all possible sets of treatment effects $\{\theta_i\}$

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

Familywise error rate

With the **familywise error rate** controlled, *for all* $\{\theta_i\}$

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

A false claim for the selected treatment occurs when $\theta_{i^*} \leq 0$ but we reject H_{i^*}.

Since this requires rejection of a true H_i, the probability of falsely claiming significance for the selected i^* is at most α.

Closed testing procedure (Marcus et al, Biometrika, 1976)

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

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

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

Closed testing procedure

The closed testing procedure controls familywise error at level α.

Proof:

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

If H_j is a true hypothesis, rejection of H_j requires rejection of $H_{\tilde{I}}$.

Thus, for any "familywise" 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 familywise error is no greater than α.

The Closure Principle

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

Phase IIb

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

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^*}: $\theta_{i^*} \leq 0$, 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 $\theta_i \leq 0$ for all $i \in I$.

The Closure Principle

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

Formally:

In order to reject H_{i^*}: $\theta_{i^*} \leq 0$, we need to reject each intersection hypothesis H_I with $i^* \in I$ at level α, based on combined Phase IIb and Phase III data.

Intuitively:

Dose i^* is chosen for the good results observed at this dose in Phase IIb.

We must adjust for this selection effect when adding the Phase IIb data on dose level i^* to the final analysis after Phase III.

Under a global null hypothesis of no treatment effect at any dose, the Phase IIb data on dose i^* should be viewed as *possibly the best results out of k ineffective doses,* rather than typical results at a single, pre-specified dose.

Intersection hypotheses

Testing an intersection hypothesis H_I: $\theta_i \leq 0$ 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

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

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: $\theta_i \leq 0$.

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) use Simes' (*Biometrika*, 1986) modification of the Bonferroni inequality:

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

Then the P-value for testing H_I is

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

Testing an intersection hypothesis

Using Simes' method:

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

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

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

The term $m\, P_{1,(j)}/j$ with $j = 1$ 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 $P_{1,I}$.

Testing an intersection hypothesis

(a) Testing H_I in Phase III

In order to reject H_{i^*}: $\theta_{i^*} \leq 0$, 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 $\hat{\theta}_{2,i^*}$ — and there is just one such test.

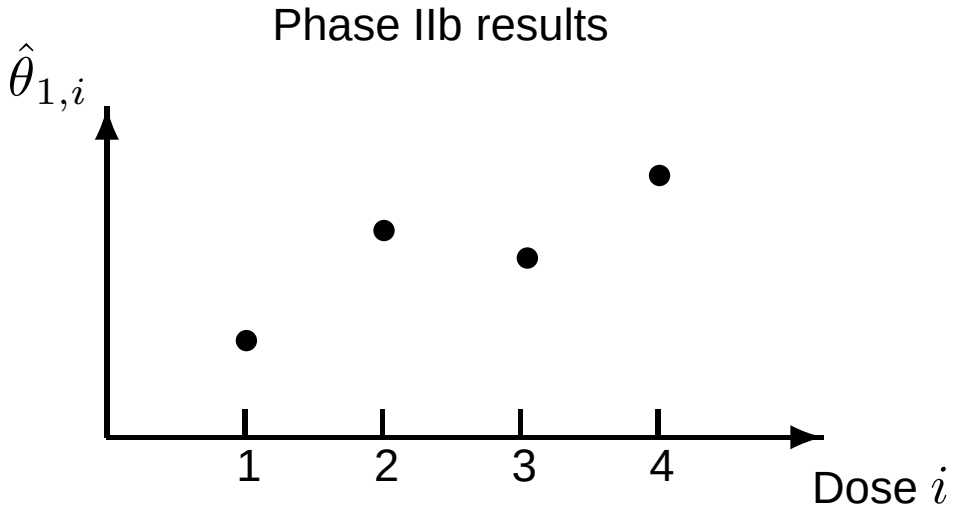
Hence, all H_I of interest have a common P-value in Phase III, $P_{2,I} = P_{2,i^*}$.

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:

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

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 $P_{1,(1)} = 0.03,\ P_{1,(2)} = 0.05,\ P_{1,(3)} = 0.2$:

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

Full workings of the Example

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

Single element sets I containing $i^* = 4$

There is just one P-value $P_{1,4} = 0.03$, so $P_{1,I} = 0.03$.

Two-element sets I containing $i^* = 4$

Consider $I = \{1, 4\}$ with ordered P-values $P_{1,(1)} = 0.03,\ P_{1,(2)} = 0.2$:

$$P_{1,I} \ = \ \min_{j=1,2} \left(2\, P_{1,(j)}/j\right) \ = \ 2 \times 0.03 \ = \ 0.06.$$

Four-element sets I containing $i^* = 4$

Ordered P-values for $I = \{1, 2, 3, 4\}$ are $P_{1,(1)} = 0.03,\ P_{1,(2)} = 0.04,$ $P_{1,(3)} = 0.05,\ P_{1,(4)} = 0.2$:

$$P_{1,I} \ = \ \min_{j=1,\ldots,4} \left(4\, P_{1,(j)}/j\right) \ = \ 4 \times 0.05/3 \ = \ 0.067.$$

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^*}: $\theta_{i^*} \leq 0$ 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 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),

Sampson and Sill (*Biometrical Journal*, 2005).

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

They have flexibility and can provide good statistical efficiency.

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

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 $k = 2$ treatments and testing *vs* control.

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

Todd & Stallard (*Drug Information Journal*, 2005) present an example where sample size per treatment is 25 in the first phase and 1400 in the second.

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

Efficiency comparisons

Jennison & Turnbull (*J. Biopharmaceutical Statistics*, 2007) consider an example with four treatments and a control. They evaluate:

- "Conventional" procedure using only second phase data in the final analysis,
- Bauer and Köhne procedures with two different combination tests,
- Thall, Simon and Ellenberg's procedure.

Sample size per treatment and control is 100 in the first phase and 500 in the second.

Procedures are defined so that the probability of proceeding to the second phase is the same in each case, so expected sample sizes are all equal.

Differences in power curves are quite slight, showing no great benefit from data combination over the two stages.

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

Other variations on the preceding methods are possible

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

2. 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.**

Selected references

Allowing *more than one treatment to progress* to Phase III:

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

Further good practices for Phase IIb/Phase III trials

Selected references, continued

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*.

There is a longstanding literature for fully sequential methods with somewhat different objectives, but the foundations of what is currently needed have been laid:

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

Robbins and Siegmund (*JASA*, 1974),

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

Further good practices for Phase IIb/Phase III trials

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.

Research agenda

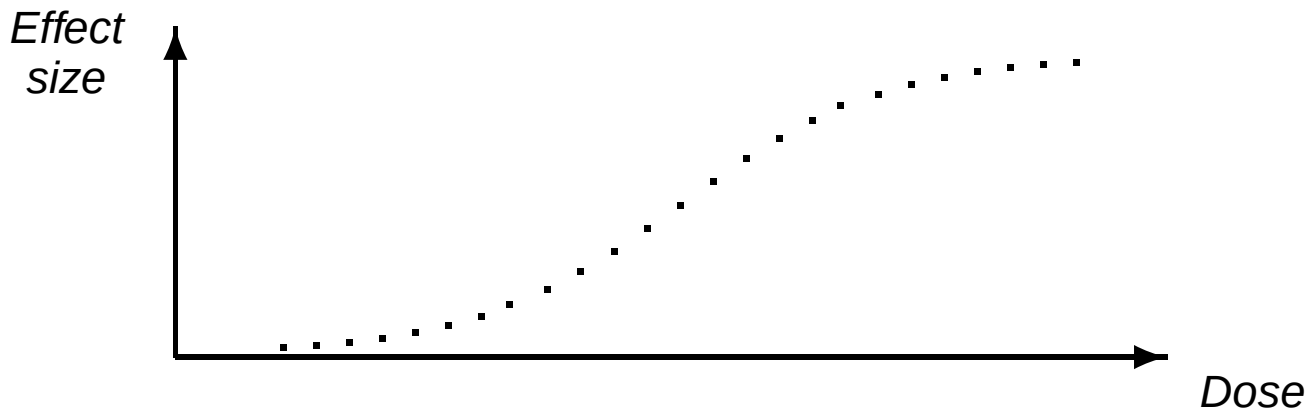
Complex calculations would be required to define such a procedure meeting specified error rates and to evaluate its properties.

Can we identify the key components needed for an overall efficient procedure?

Further good practices for Phase IIb/Phase III trials

Modelling dose response* vs *control in Phase IIb

A “nice” relationship between response and dose is usually to be expected.

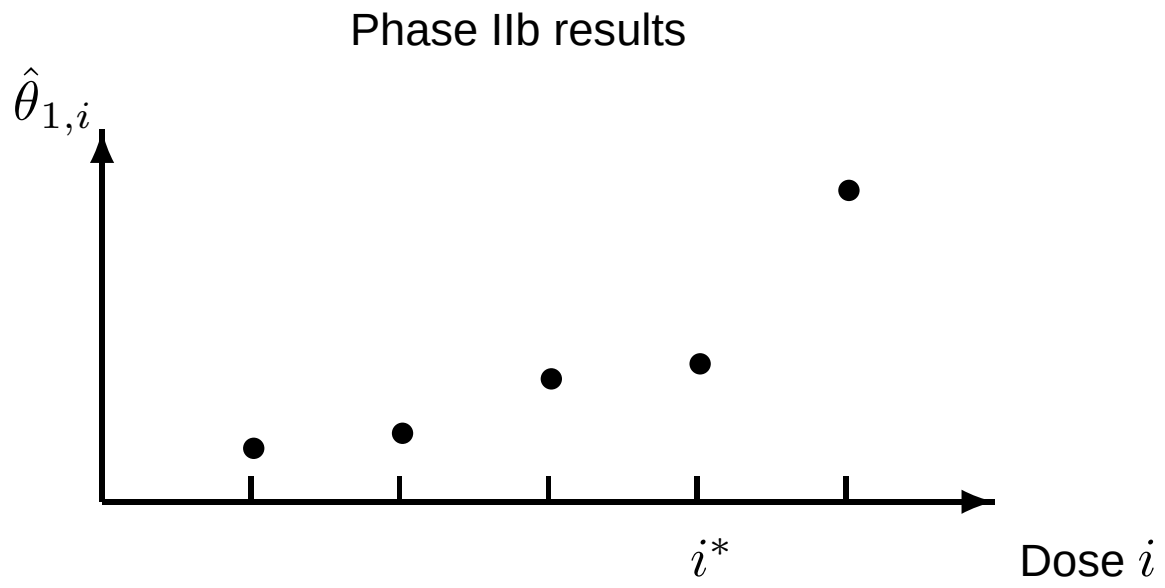


Using such a model should aid effective Phase IIb study design: see the ASTIN trial (Krams et al, *Stroke*, 2003) for an impressive example.

Integrating results from a model based Phase IIb trial into a combined analysis with Phase III data is desirable — but a model based analysis may not be viewed as providing clear cut evidence of a treatment effect at the selected dose level i^*.

Modelling dose response in Phase IIb

Previous example:



Under a typical model, any positive trend implies positive treatment effects at all dose levels — this seems to cut past the multiplicity question in Phase IIb too easily!

In the above example, the effect size at dose i^* could still be slight and the evidence for a positive effect relies heavily on the assumed model.

Showing an effect of a minimum size at dose i^* may address this issue.

Further good practices for Phase IIb/Phase III

Using a rapidly observable endpoint in Phase IIb

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

See, for example, Todd and Stallard (*Drug Information Journal*, 2005).

Research agenda

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