Futility Analysis
A Miscellany of Issues
Pharmaceutical / Publicly Funding

Professor Andy Grieve

Department of Public Health Sciences

School of Medicine

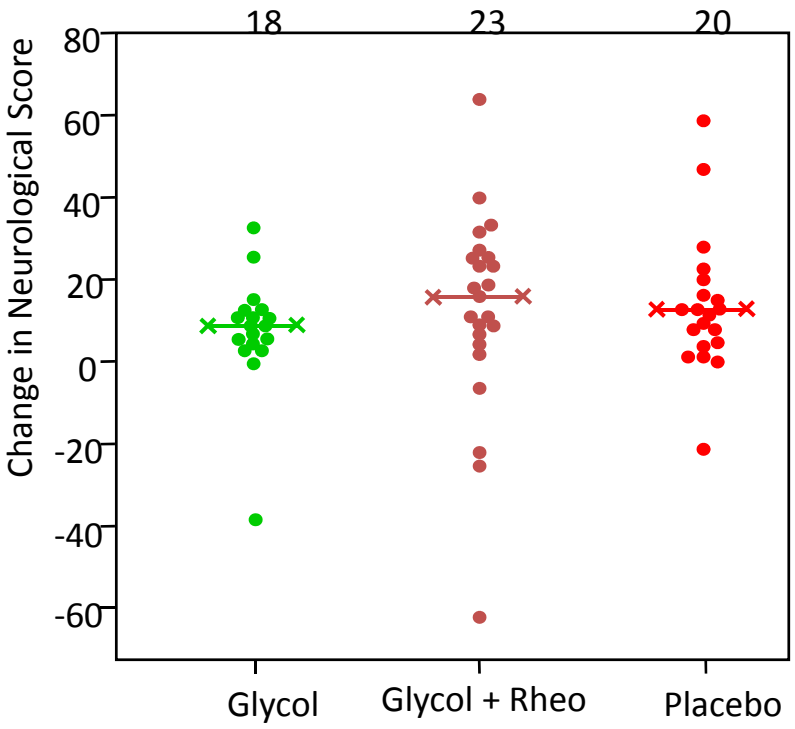
King's College London

Andy.Grieve@kcl.ac.uk

Outline

- 3 Examples of Clinical Trials That Stopped Earl
  - Confusion Between Lack-of-Benefit and Futility
- Reasons For stopping Clinical Trials in a Pharmaceutical Context
- Curtailment / Stochastic Curtailment
- Conditional & Predictive Power
  - Non-monotonicity
  - Equivalence
- Lack of Benefit of Futility?
- Arguments against the use of Predictive Power
- Publicly-funded Trials and Futility

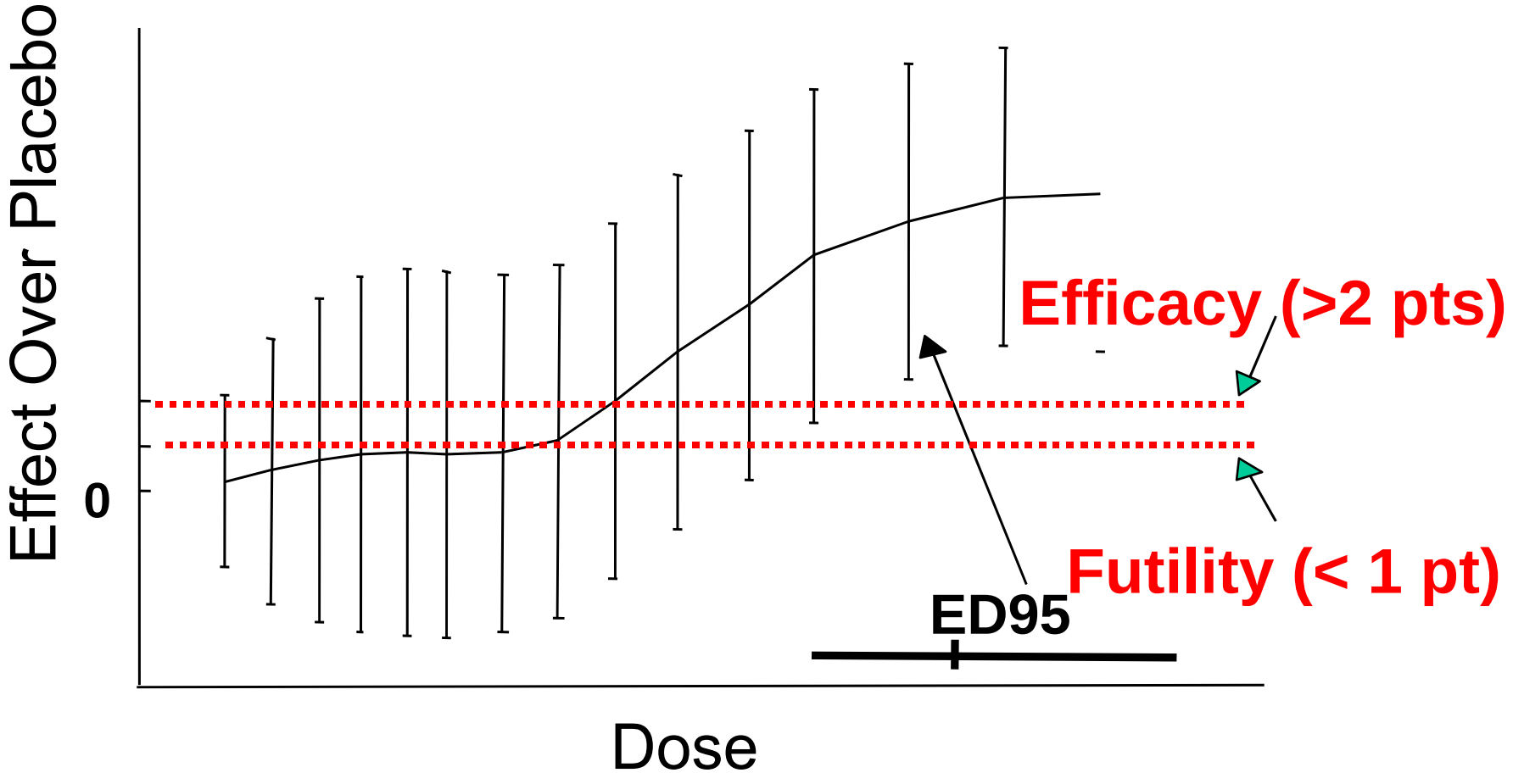
An Example of the Use of "Predictive Power" Frei et al (Stroke, 1987)



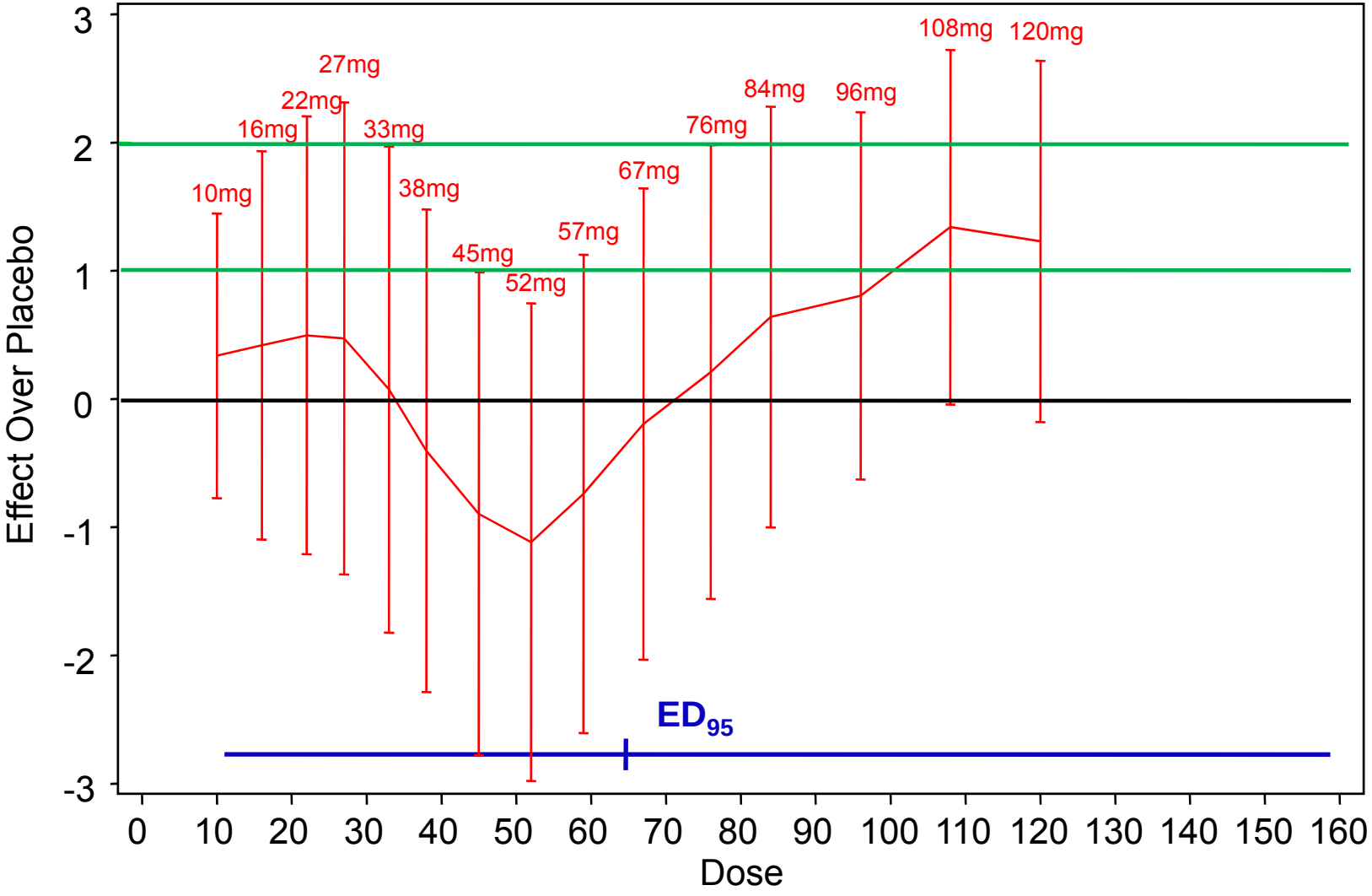
- Comparison of Glycol, Placebo and Glycol + Dextran
- Endpoint Change from Baseline Matthews Neurological Scale
- Planned : Sample size 200, interim after 100 patients
- Recruitment was slow
- Unplanned interim after 52 patients
- Predictive Probability of "achieving experimental significance with a total of 200 patients" =0.06

- STOPPED FOR FUTILITY

ASTIN Trial – Acute Stroke
FDA Workshop, Washington 2003
Dose Effect Curve



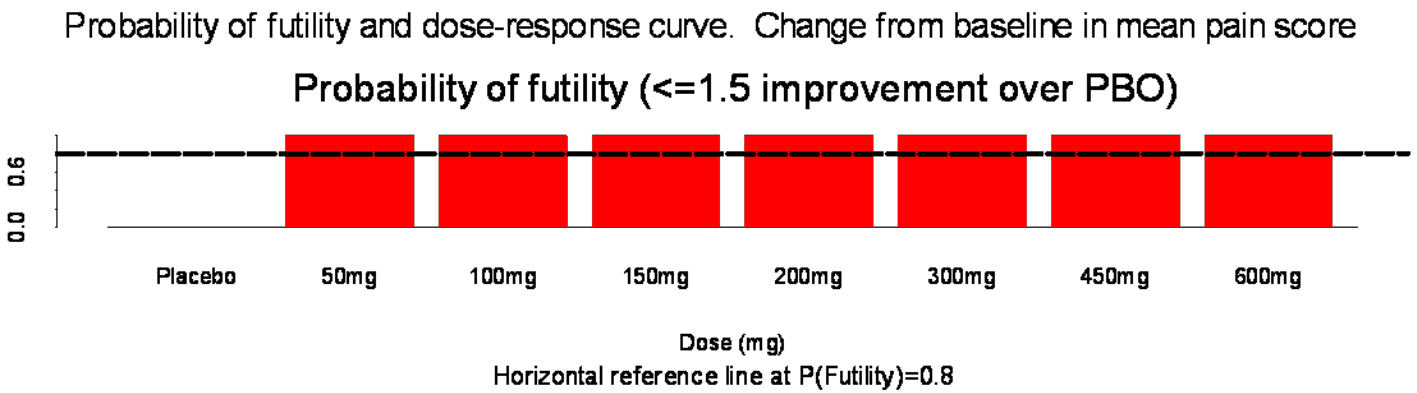
ASTIN Dose-Effect Curve + 80% Credible Intervals

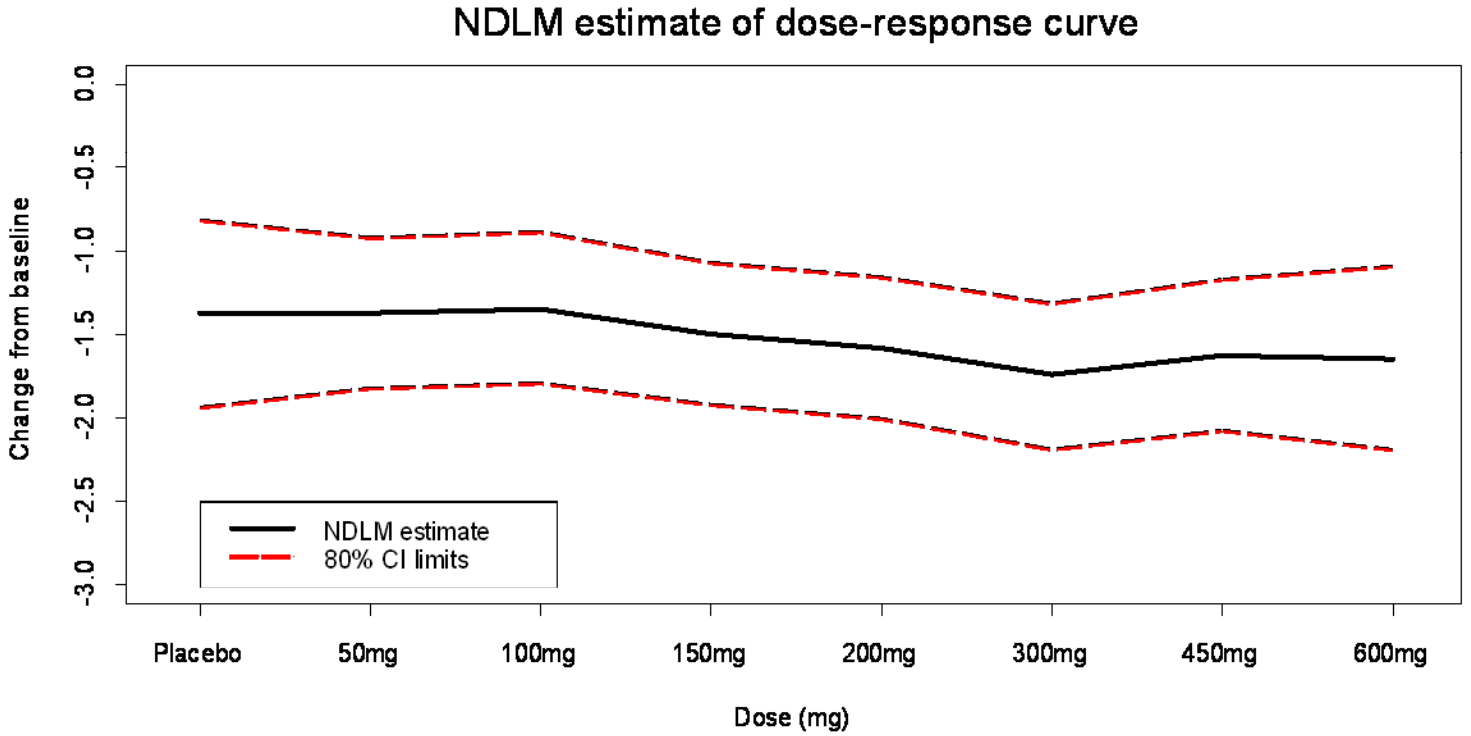


POC Study in Neuropathic Pain

Smith et al (Pharmaceutical Statistics, 2006)

Probability of futility and dose-response curve. Change from baseline in mean pain score





Reasons for Early Stopping in Pharmaceutical Sponsored Trials

- Proven efficacy - from a pharmaceutical perspective this may not be a good thing as the sponsor needs to collect enough safety information to convince regulators
- Proven safety issue(s) – of course for serious adverse events RCTs it may not be necessary to have formal safety stopping rules
- Lack of Benefit - this could be more problematic if related to purely commercial reasons
- Curtailment / Futility ?

Stopping for Commercial Reasons Commercially Futile ?

- Lievre et al (BMJ, 2001) Premature discontinuation of clinical trials for reasons not related to efficacy, safety or feasibility
- Evans and Pocock (BMJ, 2001). Editorial: Societal responsibilities of clinical trial sponsors lack of commercial pay off is not a legitimate reason for stopping a trial
- Boyd (BMJ, 2001). Commentary: Early discontinuation violates Helsinki principles.
- Cannistra (J Clin Oncol, 2004). The ethics of early stopping rules: Who is protecting whom?
- Psaty and Rennie (JAMA, 2003). Stopping medical research to save money – A broken pact with researchers and patients.
- Iltis (J Med ethics, 2004). Stopping trials early for commercial reasons: the risk-benefit relationship as a moral compass.
- Trotta et al (Ann. Oncology, 2008). Stopping a trial early in oncology; for patients or for industry?

Illustration of Curtailment in a Clinical Trial

- Two treatments, 20 patients per group
- After 10 patients per group following results :

- Active 4/10 , Control 8/10 : minimum possible control response at completion is 8/20
- Only active response rates which are significant given 8/20 in controls are :
- 15/20 , 16/20 , 17/20 , 18/20 , 19/20 , 20/20 - Impossible !!
- Of course if the active had been 5/10 then – in theory a significant result could have been possible – but how likely is it to get 10/10 on active and 0/10 on controls.
- We need to be able to calculate the appropriate probability – but under what assumption?

Stochastic Curtailment
Conditional and Predictive Power

- Assume m pts are treated in each of 2 groups and that the posterior distribution of δ, the difference in means, is $N(y_m, \sigma_\delta^2/m)$ where y_m is the difference in sample means and $\sigma_\delta^2 = 2\sigma^2$.
- The posterior probability that δ is positive is

$$P_m = P(\delta > 0 \mid y_m) = \sqrt{\frac{m}{2\pi\sigma_\delta^2}} \int_0^\infty \exp\left[\frac{-m}{2\sigma_\delta^2}(\delta - y_m)^2\right] d\delta = \Phi\left(\frac{m^{1/2} y_m}{\sigma_\delta}\right) \quad \text{(1-sided p-value)}$$

- By analogy the posterior probability after N=m+n pts in each group is

$$P_N = \Phi\left(\frac{N^{1/2} y_N}{\sigma_\delta}\right)$$

where $y_N = \dfrac{m y_m + n y_n}{N}$ and y_n is the difference in means based on a further 2n pts

Stochastic Curtailment Conditional and Predictive Power

- If a trial is regarded as a success when $P_N > 1-\varepsilon$ implies

$$y_n > \frac{N^{1/2}\sigma_\delta z_\varepsilon - m y_m}{n}$$

- By definition, $p(y_n \mid \delta) = \sqrt{\frac{n}{2\pi\sigma_\delta^2}} \exp\left[\frac{-n}{2\sigma_\delta^2}(y_n - \delta)^2\right]$ conditional on δ - the planned alternative - so that

$$CP = \Pr\left(y_n > \frac{N^{1/2}\sigma_\delta z_\varepsilon - m y_m}{n} \middle| \delta \right) = \Phi\left[\sqrt{\frac{N}{n}} z_\varepsilon + \frac{m y_m + n\delta}{\sqrt{n}\sigma_\delta} \right]$$

Stochastic Curtailment

- A criticism of the conditional power argument is that the calculations may be based on values of θ that have no support from the data.
- This can be simply overcome by utilising a Bayesian/Predictive approach.
- Here instead of conditioning on the original alternative, the conditioning is based on what has been observed.

Stochastic Curtailment Conditional and Predictive Power

- The predictive distribution for y_n conditional on y_m is:

$$p(y_n \mid y_m) = \sqrt{\frac{nm}{2\pi N\sigma_\delta^2}} \exp\left[\frac{-nm}{2N\sigma_\delta^2}(y_n - y_m)^2\right]$$

so that

$$PP = \Pr\left(y_n > \frac{-N^{1/2}\sigma_\delta z_\varepsilon - my_m}{n} \middle| y_m\right) = \Phi\left[\sqrt{\frac{m}{n}} z_\varepsilon + \sqrt{\frac{mN}{n}} \frac{y_m}{\sigma_\delta}\right]$$

$$= \int_\delta \Phi\left[\sqrt{\frac{N}{n}} z_\varepsilon + \frac{my_m + n\delta}{\sqrt{n}\sigma_\delta}\right] p(\delta \mid y_m) d\delta$$

- This is not just a Bayesian result since by definition

$$y_n - y_m \sim N(0, \sigma_\delta^2(1/m + 1/n)) \Rightarrow y_n \sim N(y_m, \sigma_\delta^2(1/m + 1/n))$$

Data Driven Rules

- A stopping rule based on

$$P_m = P(\delta > 0 \mid y_m) = \sqrt{\frac{m}{2\pi\sigma_\delta^2}} \int_0^\infty \exp\left[\frac{-m}{2\sigma_\delta^2}(\delta - y_m)^2\right] d\delta = \Phi\left(\frac{m^{1/2} y_m}{\sigma_\delta}\right)$$

posterior probability of a positive treatment effect at the interim or on

$$PP = Pr\left(y_n > \frac{-N^{1/2}\sigma_\delta z_\varepsilon - m y_m}{n} \middle| y_m\right) = \Phi\left[\sqrt{\frac{m}{n}} z_\varepsilon + \sqrt{\frac{mN}{n}} \frac{y_m}{\sigma_\delta}\right]$$

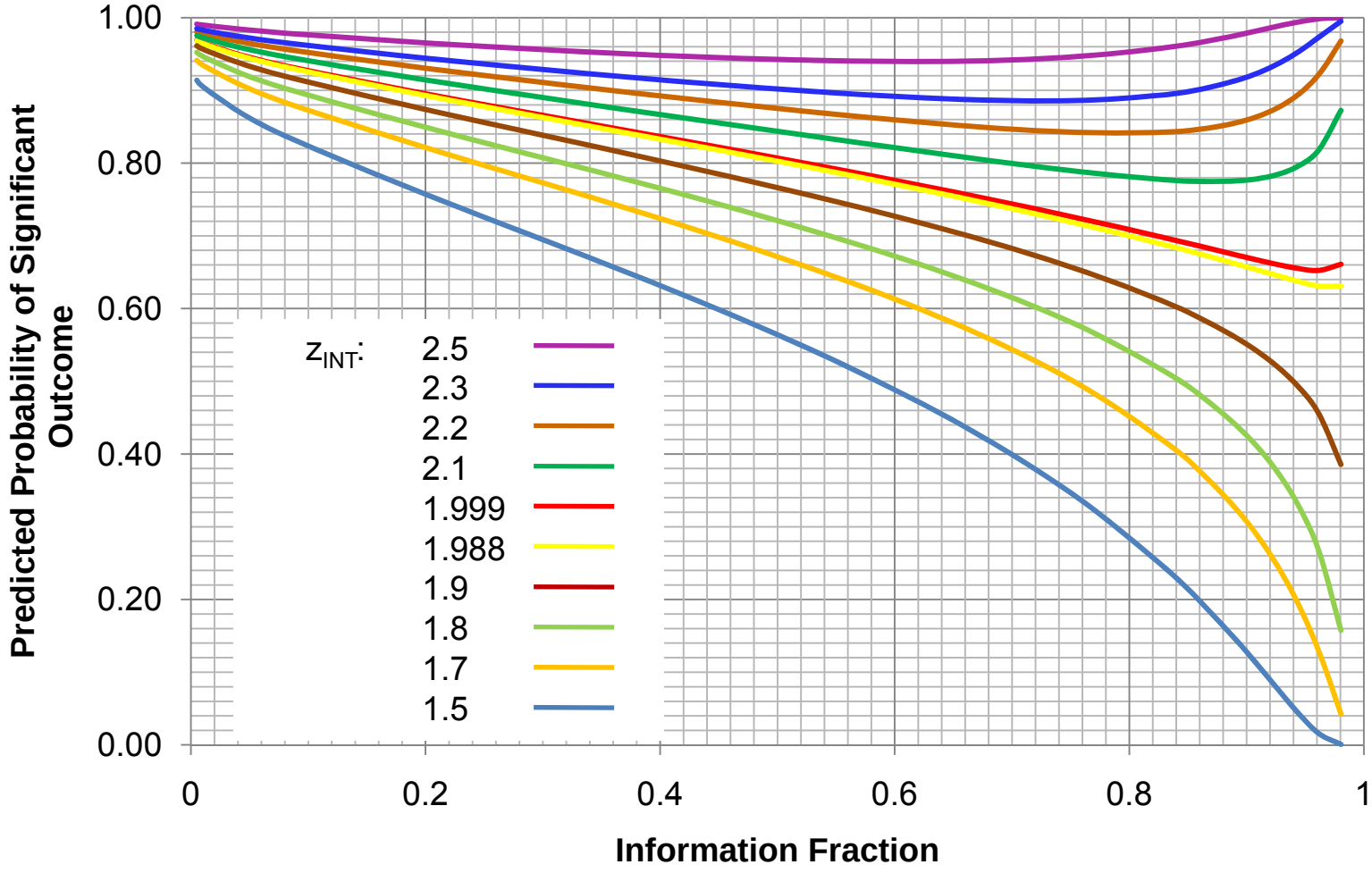
both depend only on the interim data y_m and an equivalence can be found between them

Predictive Power

- One criticism of predictive power is that it can be non-monotonic as a function of f the information fraction):
  - Predictive Power: $PP = \Phi\left(\frac{z_{INT} + \sqrt{f} z_{\varepsilon}}{\sqrt{1-f}}\right)$
  - Minimum at : $\sqrt{f} = \frac{-z_{\varepsilon}}{z_{INT}}$

Only if $z_{INT} > z_{\varepsilon}$ otherwise monotonic (not quite true because in terms of n and m, f is not continuous.)

Conditional / Predictive Power

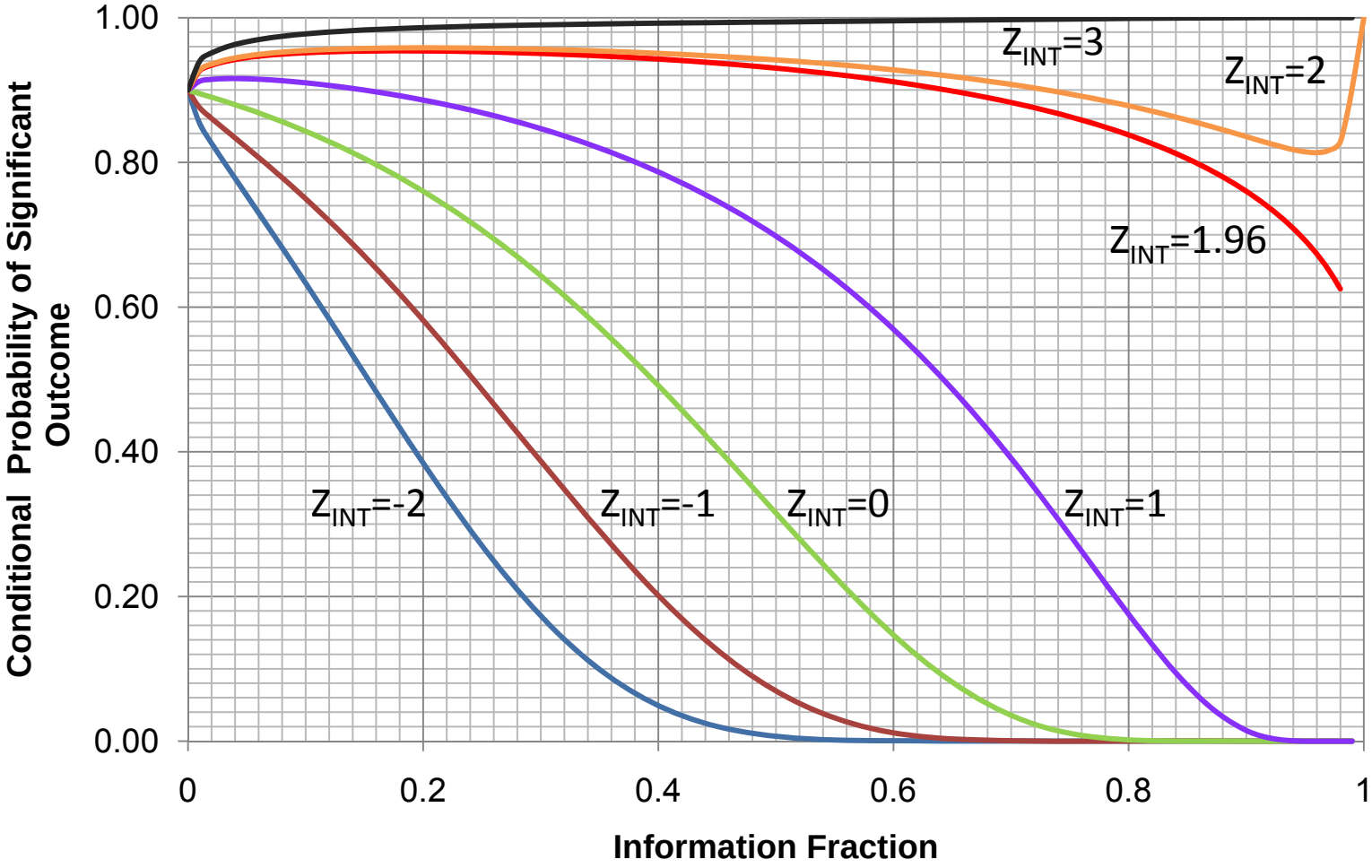


Predictive Power

- Should we be worried about non-monotonicity?
- Not overly
- It only occurs when z_{INT} is beyond the critical value at the end of the trial. In other words when there is "unadjusted" evidence of a treatment effect. From a pure Bayesian/likelihood perspective this is already sufficient evidence to stop the trial.

Conditional Power has the Same Defect

Proschan, Lan & Wittes, “Statistical Monitoring of Clinical Trials “, 2006



Futility or Lack of (sufficient) Benefit

- What is futility?
- In my mind futility is a prospective/predictive concept
- I remain to be convinced that early stopping should be based on a prediction.

Aside on Early Stopping for Lack of Benefit

- I have been involved in 10 adaptive clinical trials – either as designer, or as a member of a DSMB
- ALL have stopped early for lack of Benefit / Futility
- I'm not surprised / I'm pleased
- > 95% of all chemical / biological considered as medicines fail
- Between 80 and 90% fail in phases I-III
- I therefore have a high subjective probability on starting a study that the drug doesn't work

Arguments Against Predictive Power

- Bayesians have criticised p-values as being probabilities of events that "could have happened but didn't"
- Predictive power is a probability of events that "might happen but haven't - yet".

- Armitage (Cont Clin Trials, 1991) argues against the use of predictive power as a formal stopping rule – as do Spiegelhalter, Abrams and Myles

  - It gives undue weight to "significance"
  - Makes strong assumptions about the comparability of future data with the past – for example if future data involve follow-up there may be a reliance on an assumption of proportional hazards

Publicly Funded Trials and Futility

- Should futility/lack of benefit be used in publicly funded trials?
- In some cases yes.
- For example, I see no scientific reason why futility / lack of benefit should not be used in experimental medicine studies
- The non-scientific reason might have to do with the appointment of RAs, post docs etc as part of the grant