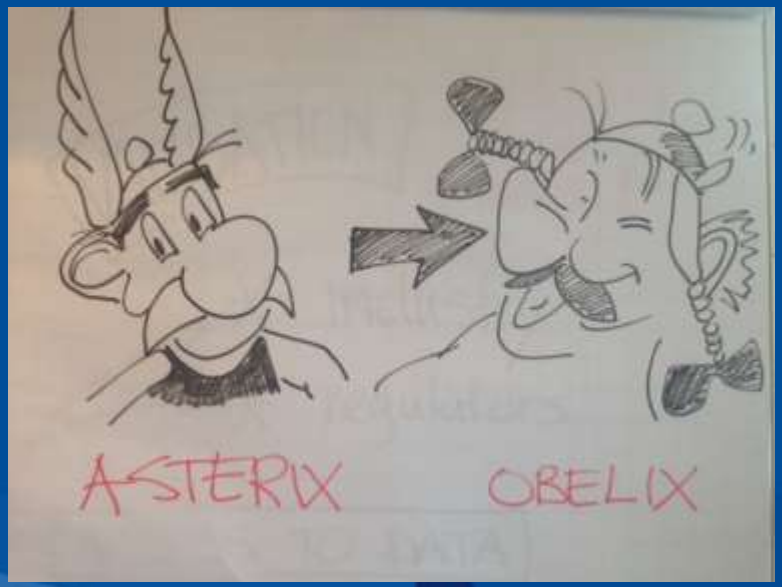
UMC Utrecht

Hybrid Bayesian-frequentist approaches for small sample trial design: examples and discussion on concepts.

Stavros Nikolakopoulos
Kit Roes
UMC Utrecht



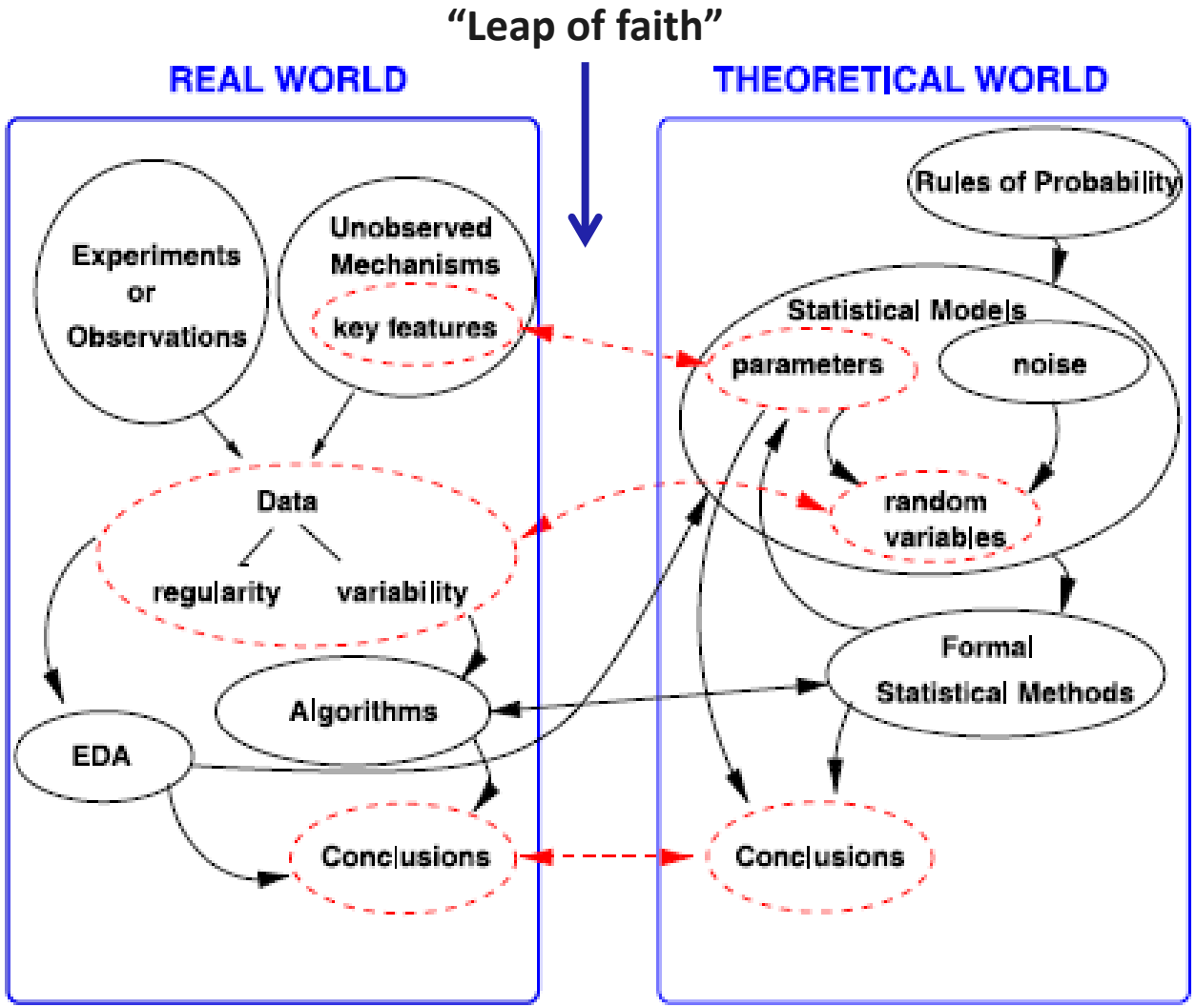
Outline

- Comfortable or not with hybrid Bayesian-frequentist approaches.
- Motivation to explore these models
- Prospectively planned (dis)counting of prior data
- For discussion……

Comfortable or not

- "Controversies in the field of mathematical statistics seem largely to have arisen because statisticians have been unable to agree upon how theory is to provide, in terms of probability statements, the numerical measures most helpful to those who have to draw conclusions from observational data." E.S. PEARSON (1955)

Comfortable or not



Robert E Kass. Statistical Inference: The Big Picture. *Stat Scie*, 2011

Motivation to explore these models
(for small populations)

asterix

- Probability statements about parameters ✓ ✗
- Adequately accounting of uncertainties ✓
- Prediction of future observations ✓
  (incl predictive/conditional power)
- **Incorporate prior data (information) model based** ✓
  (to increase sensitivity)
- Incorporate prior belief (in decision making context) ✗

Prospectively planned (dis)counting of prior data

asterix

Illustrative example: Clinical trial in pediatric surgery/pain.

Isobaric bupivacaine (**C**ontrol) vs bupivacaine+clonidine 2g/kg (**T**reatment) in adolescents (age 10-15).

Primary outcome: the mean duration of sensory block (minutes).

Sample size based on previous study with 21 patients per group, standardized effect size of 0.76.

37 patients per group for 90% power with a two-sided of 5%.

Some differences between studies, considered of no impact – hence pooling possible.

Kaabachi, A et al. (2007), Anesth. & Analg.105(2):516–519.

Prospectively planned (dis)counting of prior data

Setting of:

- Small populations – rare diseases.
- Development of new treatments.
- Design/analysis of new ("Phase III") study (D_1), with prior availability of a small earlier study (D_0).

Aim:

- Can we strengthen evidence from D_1 by prospectively defined pooling.
- Controlling frequentist properties (type 1 error).
- Considering potential heterogeneity.

Prospectively planned (dis)counting of prior data

So:

- Include first study as “prior information” into analysis of the second (Phase III).
- Weight of study decreases with increasing heterogeneity (in some sense).
- Assess and control type 1 error properties (under classical data generating mechanism).

Prospectively planned (dis)counting of prior data

Concept of power priors.

$$\pi(\theta|D_0,\gamma) \propto L(\theta|D_0)^{\gamma}\pi_0(\theta)$$

where:

D_0 represents the data of the first small trial.
$\pi_0(\vartheta)$ the general (flat) prior.
$\pi(\vartheta \mid D_0, \gamma)$ the posterior to be used as prior to the new trial.

$\gamma \in [0,1]$ defines the level of (down) weighting the data.

Power priors

$$\pi(\theta|D_0,\gamma) \propto L(\theta|D_0)^{\gamma}\pi_0(\theta)$$

γ ∈ [0,1]:

- (Assumed) known & fixed.
- Assigned a prior distribution.
- *Estimated*.

New concept:

- Level of down-weighting prior evidence depends on (dis)similarity between prior and current data.
- Control frequentist properties.

Competing concept (based on power priors):

- Test-then-Pool (TtP). Test $H_0 : \mu_T = \mu_0$ and pool conditionally.

Calibrated power prior

Prior for the new data based on n_0 observations of the small previous trial.

$$\mu|\sigma^2 \sim N(\mu_0, \frac{\sigma^2}{n_0})$$

Estimate:

$$\hat{\gamma} = \begin{cases} \frac{\sigma^2}{n_0} \Big/ \left[\left(\frac{\bar{X}-\mu_0}{z_{1-c/2}} \right)^2 - \frac{\sigma^2}{n_1} \right], & \text{if } \bar{X} < \mu_0 + z_{c/2}\sigma_{pr} \ \vee \ \bar{X} > \mu_0 + z_{1-c/2}\sigma_{pr} \\ 1, & \text{if } \mu_0 + z_{c/2}\sigma_{pr} \leq \bar{X} \leq \mu_0 + z_{1-c/2}\sigma_{pr}. \end{cases}$$

Positive decision at the end of the trial if:

$$Pr(\mu > 0|D_1) > \eta$$

To assess frequentist properties, a fixed data generating mechanism is assumed (with mean μ_T)

Calibrated power prior

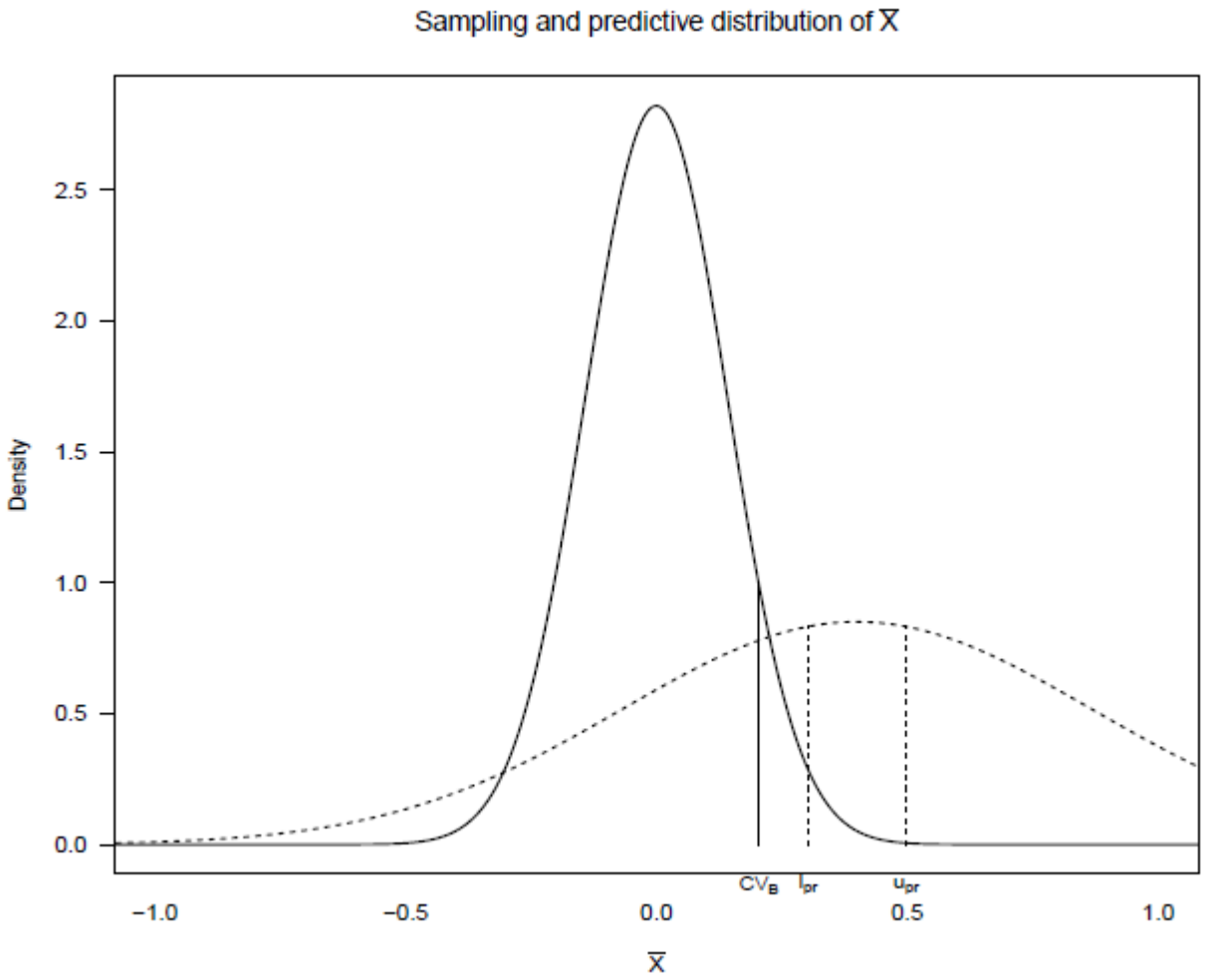


Figure 1: Sampling (solid line) and predictive (dashed line) distributions of $\bar{X}$ for $\mu_T = 0, \mu_0 = 0.4, \sigma^2 = 1, \eta = 0.95, n_0 = 5, n_1 = 50$ and $z_{c/2} \approx 0.2$ so $c \approx 0.84$.

Calibrated power prior: Type 1 error

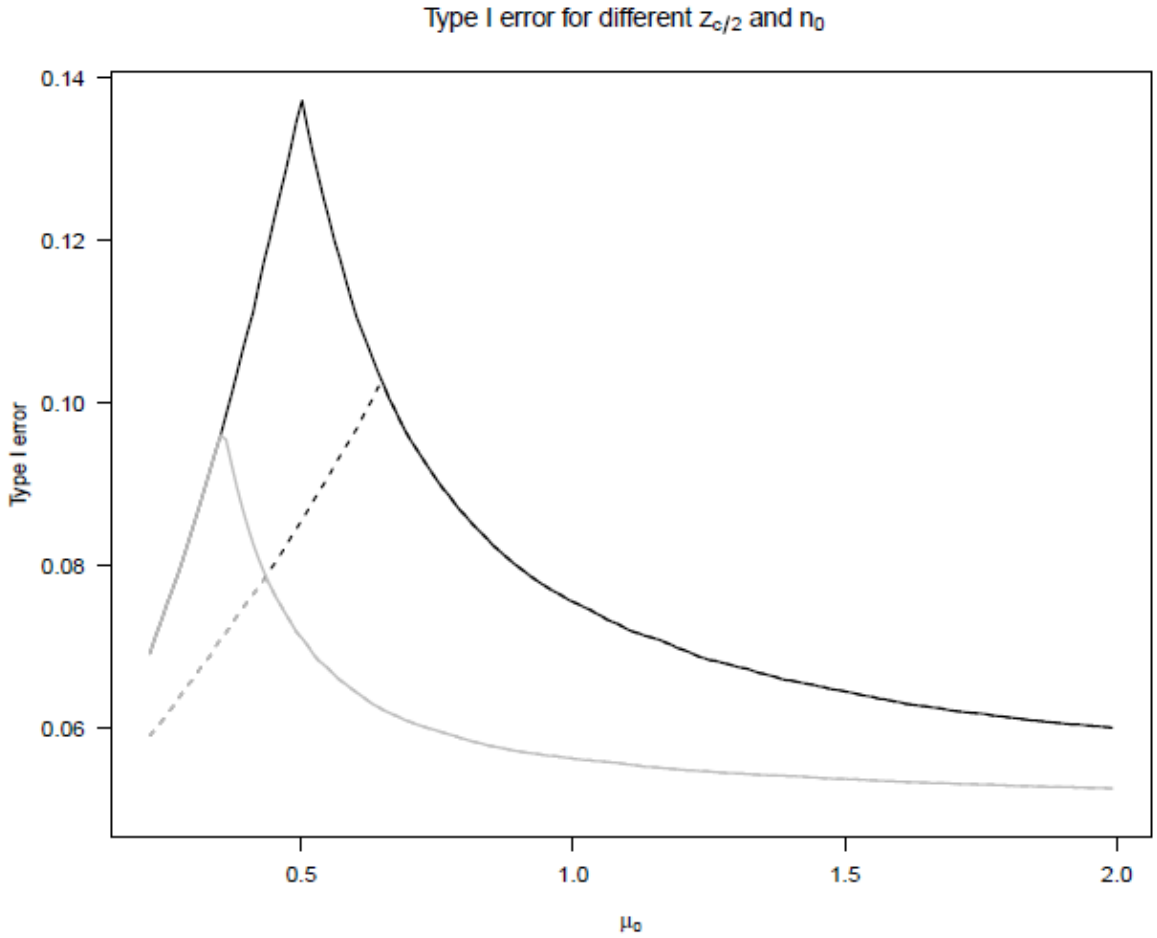


Figure 5.2: Type I error for $PDCCPP$ when $n_0 = 10$ (solid lines) and $n_0 = 5$ (dashed lines) for $z_{1-c/2} = 1$ (black lines) and for $z_{1-c/2} = 0.5$ (grey lines) as a function of μ_0; $\sigma^2 = 1, \eta = 0.95$ and $n_1 = 50$.

Calibrated power prior: MSE

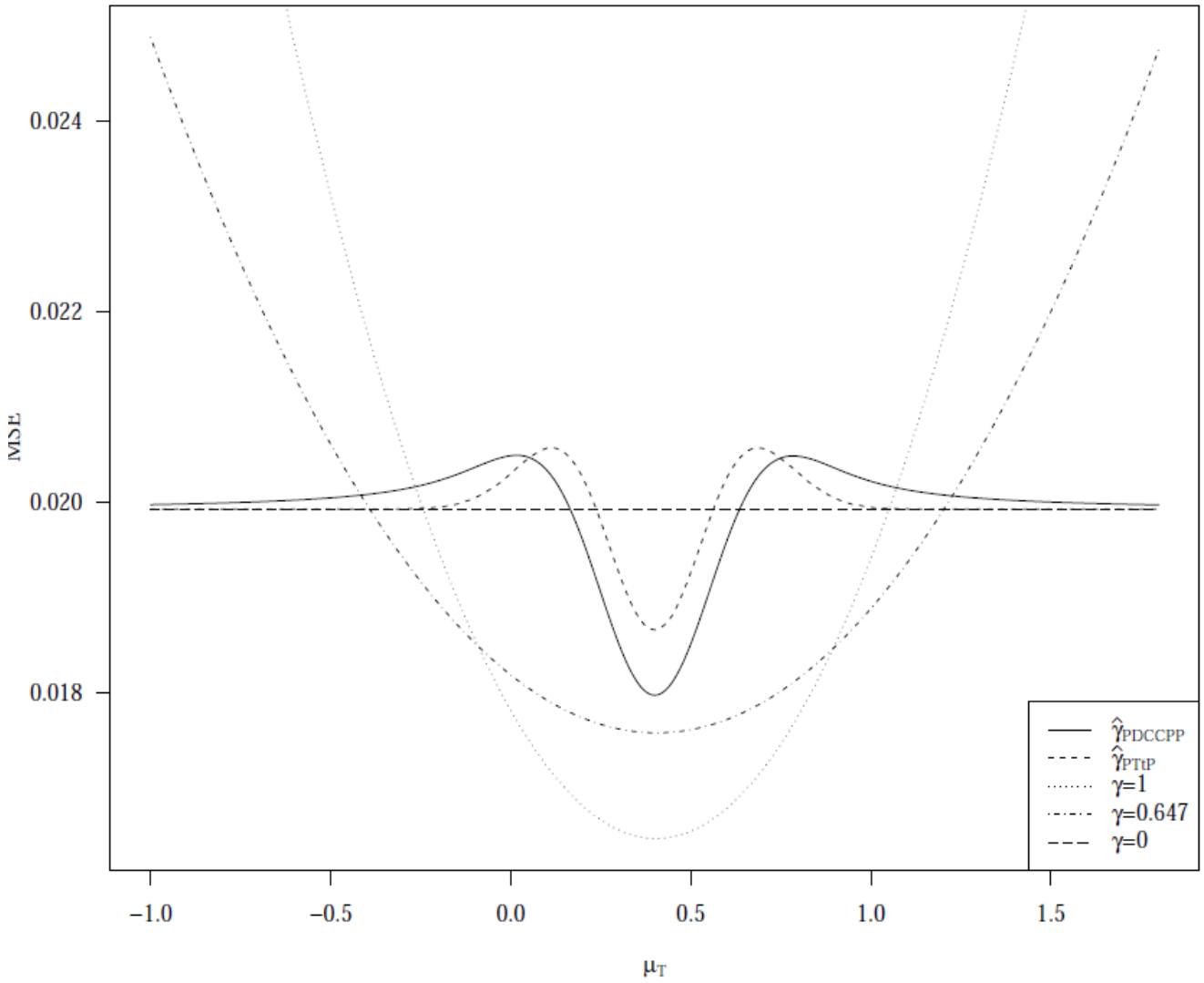


Figure 5.4: MSE for different values of γ, as a function of μ_T. The $PDCCPP$ and $PTtP$ estimates are calibrated to have a type I error of 6.5%, and so is the fixed γ of 0.647; $\mu_0 = .4$, $\sigma^2 = 1, \eta = 0.95$ $n_0 = 5$ and $n_1 = 50$.

Example continued

Recall:

Isobaric bupivacaine (**C**ontrol) vs bupivacaine+clonidine 2g/kg (**T**reatment)

Sample size based on previous study with 21 patients per group, standardized effect size of **0.76**: $\mathbf{D_0}$

New trial: 41 patients per group.

Result: Standardized effect size of **0.58**: $\mathbf{D_1}$

Example continued

	Type 1 error (one-sided)	
	0.035	0.040
z-value	0.385	0.475
No borrowing		
$\delta_1 \mid \gamma = 0$	0.580	
$P(\delta > 0 \mid \gamma = 0)$	0.996	
95% CrI $\mid \gamma = 0$	(0.28, 0.88)	
Calibrated power prior		
γ	0.557	0.992
$\delta_1 \mid \gamma$	0.619	0.640
$P(\delta > 0 \mid \gamma)$	0.999	0.999
95% CrI $\mid \gamma)$	(0.35, 0.88)	(0.39, 0.89)
Full borrowing		
$\delta_1 \mid \gamma = 1$	0.640	
$P(\delta > 0 \mid \gamma = 1)$	0.999	
95% CrI $\mid \gamma = 1$	(0.39, 0.89)	

Discusion

1. Prospectively defined weighting is attractive.

2. Clear link with (cumulative) meta-analysis & heterogeneity.

3. Not sure if problem formulation is optimal / leads to optimal solutions.

4. Conceptually it is hard to consider it (truly) Bayesian.

5. Any suggestions for further development are welcome!