VIII BIAS Annual Congress, Verona 30/06/2016

Bayesian methods for the DMC

Mauro Gasparini, with thanks to **Gaëlle Saint-Hilary**
Dipartimento di Scienze Matematiche "G.L. Lagrange"
Politecnico di Torino
mauro.gasparini@polito.it
http://calvino.polito.it/~gasparini

This project has received funding from BIAS and from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567.

My current Ph.d. students

Gaëlle Saint-Hilary, sponsored by Servier, on Quantitative Decision Making in Drug Development.

José Jimenez, Marie Curie student sponsored by the EU IDEAS consortium*, on Innovative designs for combination of existing therapies.

Elvira Erhardt, Marie Curie student sponsored by the EU IDEAS consortium*, on Modelling and simulation for the early development of a modified administration route (with ABC Farmaceutici).

Maryam Zolghadr, visiting Astrazeneca Göteborg, on surrogate markers in respiratory drug development.

Plus Luca Grassano, a laurea magistrale student doing his internship at Roche Paris.

The Statistician in a DMC

DMC is one of the safety activities in a clinical trial related to an investigational drug.

A statistician participating to a DMC is mostly responsible for monitoring the statistical analysis of safety data.

The approach is a supervising one, more qualitative than technical, since many analysis activities (such as interim analysis) will be performed by the sponsor according to the protocol.

Some Bayesian ideas, rather than formal methods, will be reviewed in this talk.

Stopping for efficacy at Interim Analysis

We start with efficacy monitoring, since most of methodology is about that.

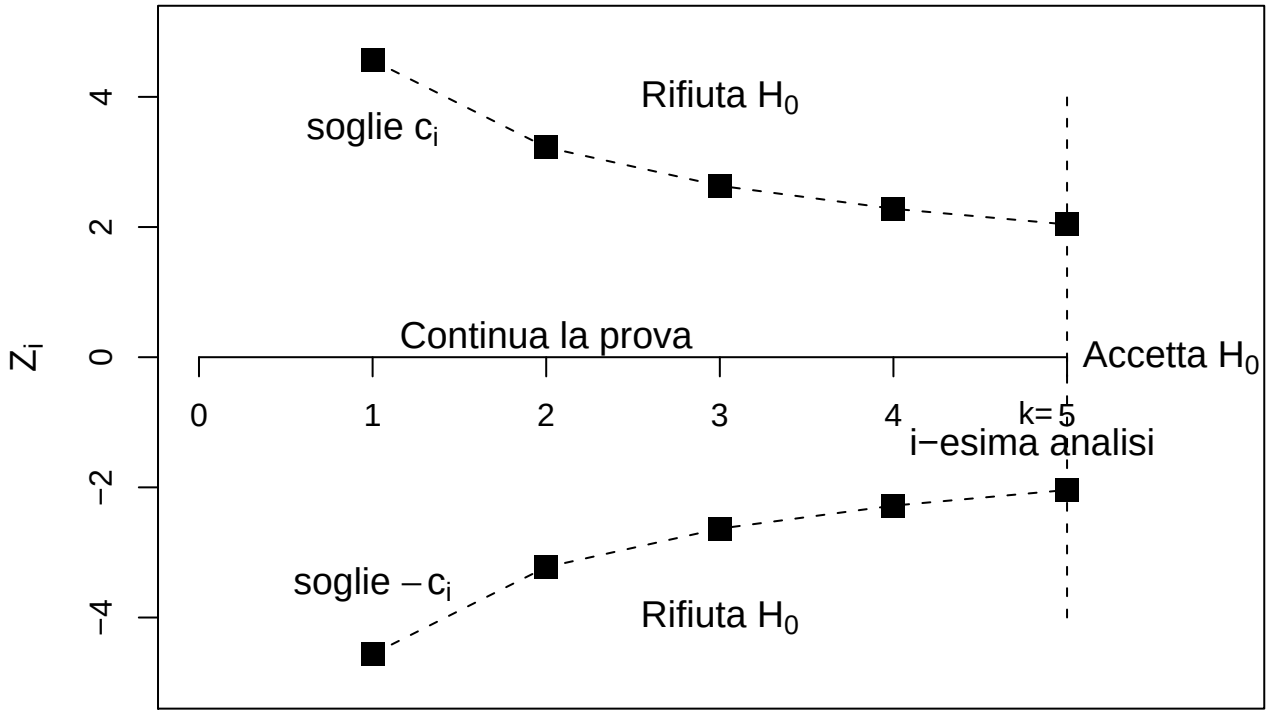
Frequentist (non-Bayesian) methods for interim analysis are based on Group Sequential Methods (GSM, e.g. Pocock, O'Brien and Fleming, alpha-spending) and stochastic curtailment (conditional power): mostly geared at efficacy. See the classic Jennison and Turnbull (2000).

Early stop can be justified:

- for futility, when there is little hope to demonstrate efficacy (differential effect of treatment)
- for utility (success), when there is overwhelming evidence in favor of differential effect of treatment

In GSM, the main concerns is multiplicity, since type I error is the main concern of authorities.

Grafico di O'Brien e Fleming con $k = 5$ sguardi



Bayesian Idea # 1: use a community of priors

Monitoring for efficacy can also be done in Bayesian terms.

A popular idea from Bayesian Robustness is to use a community of priors. In clinical trials, the idea takes a very practical form of (at least) three priors to deal with (Spiegelhalter, Abrams and Myles, 2004, SAM 2004 from now on):

- a sceptical prior
- a moderate prior (sometimes omitted)
- an optimistic prior

A DMC statistician may focus on one of them, depending on the issue at stake. For example:

- to stop for futility, temper an enthusiastic prior
- to stop for efficacy, convince a sceptical prior

Bayesian Idea # 2: separate demands and beliefs

The Bayesian approach allows for "...distinction between *demands*, as expressed in their range of equivalence, and their *expectations* or *beliefs*, as represented by the prior distribution." (SAM 2004)

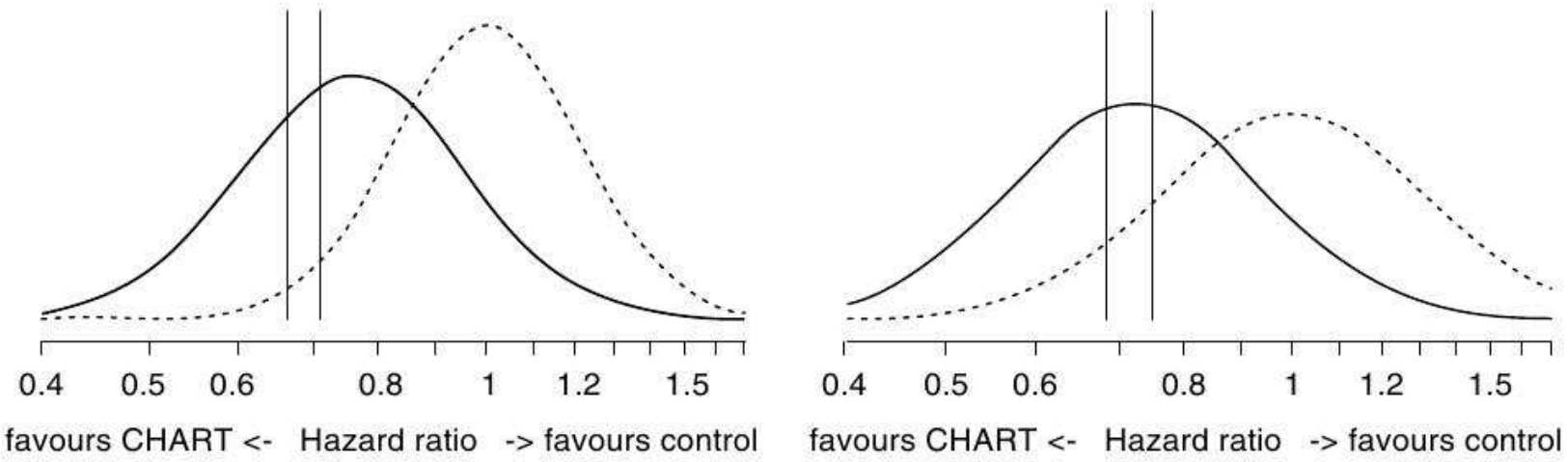


Figure 6.2 Clinical and sceptical priors superimposed on an assessed average clinical range of equivalence. Probabilities of lying below, within and above the range of equivalence are given both for clinical and sceptical priors. The juxtaposition of the clinical priors and ranges of equivalence suggests a reasonable basis for randomisation.

Example # 1: Bayesian Monitoring in the CHART trial

From SAM 2004, building on previous work.

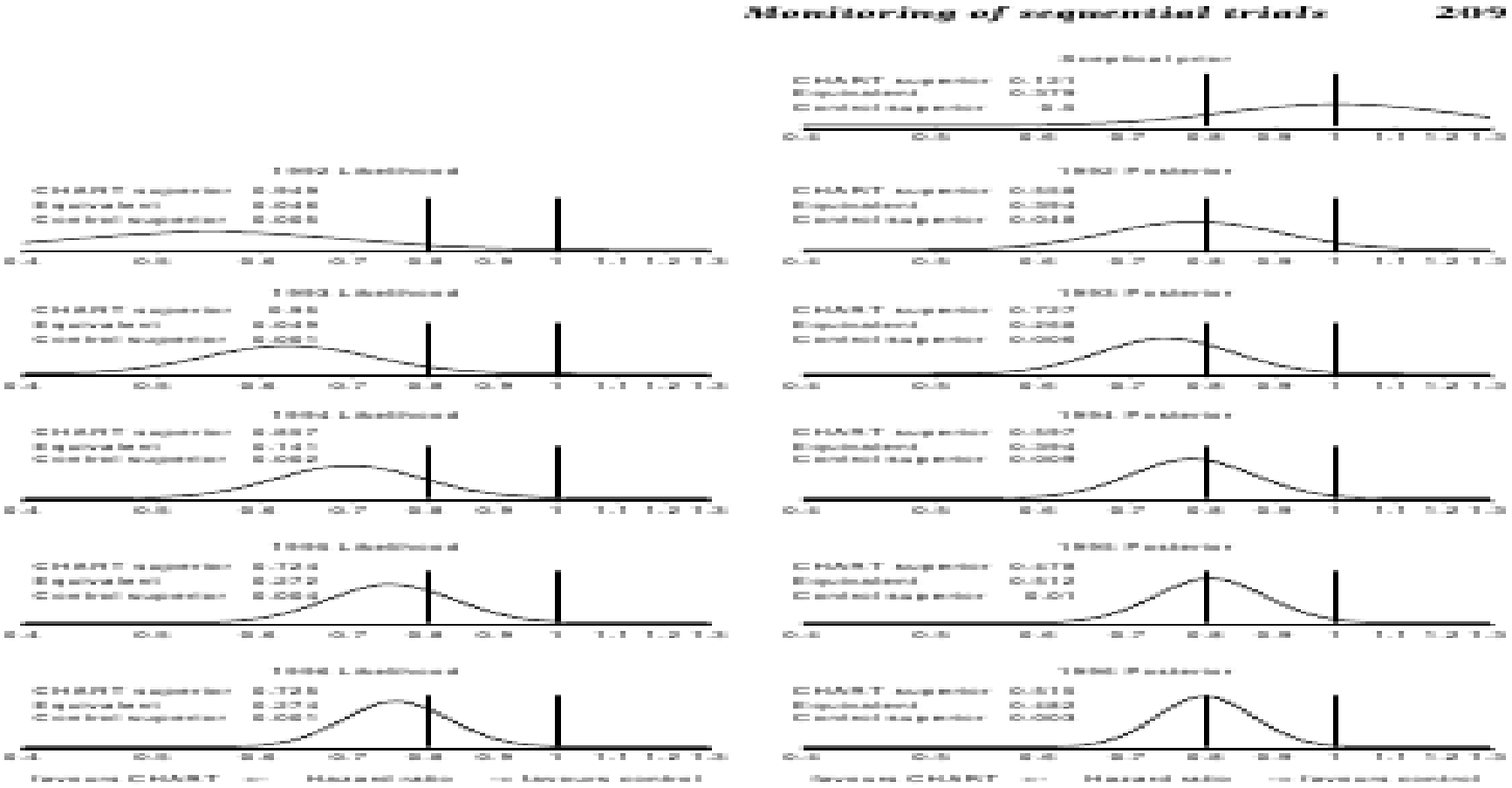


Figure 6.7 Prior, likelihood and posterior distributions for the CHART lung cancer trial assuming a sceptical prior. The likelihood becomes gradually less extreme, providing a very stable posterior estimate of the treatment effect when adopting a sceptical prior centred on a hazard ratio of 1. Demands are based on a 7% improvement from 15% to 22% 2-year survival, representing a hazard ratio of 0.80.

Example # 2: double criteria designs

Two simultaneous requirements in terms of posterior probabilities especially developed at Novartis.

In a noninferiority trial, for example, where PFS is the endpoint, require:

- P($HR \leq 1.1$|data) ≥ 0.5
- P($HR \leq 1.27$|data) ≥ 0.9

Neuenschwander *et al.* (2011) to accompany Motzer *et al.* (2014), JCO 32, 2765–2772.

Stopping for safety at Interim Analysis

Sometimes a trial may be stopped for safety concern, when the treatment may cause too many adverse events (AE, as coded in MedRA).

Type-II error here (non seeing important safety problems) is more important than type-I (false positive safety problems). But a similar multiplicity problem arises, which we now see how to control in Bayesian terms.

Here the DMC may have a role in recommending stopping for safety.

Monitoring often happens with blinded data, since often if there are safety problems they are due to treatment.

Bayesian flexibility for the DMC

According to Bayesian ideas, a DMC is flexible on at least two grounds (SAM 2004, explicit sec. 6.6.6 on Bayes and DMC):

- incorporating external information to make decisions and update priors
- make recommendations on subgroups (or arms) of patients

Example # 3: subpopulation trials in the BELLE-4 trial

Is the therapy working in the full population or only in a subpopulation? (This example is efficacy, but similar ideas work for safety).

Bayesian criterion to select populations in which the therapy is more effective (targeted therapies): "Decision rules based on predefined thresholds of PFS predictive probabilities of success in the full and PI3K pathway-activated populations were used to decide whether the study would enter Phase III (in the full population or restricted to the PI3K pathway-activated population), or be stopped for futility." Martin et al. (2015): `http://mct.aacrjournals.org/content/14/12_Supplement_2/A166`

Based on methodology work by Brannath et al (2009).

Safety Data Analysis

The statistical analysis of safety data concerns:

- adverse events (in terms of reported symptoms)
- laboratory data

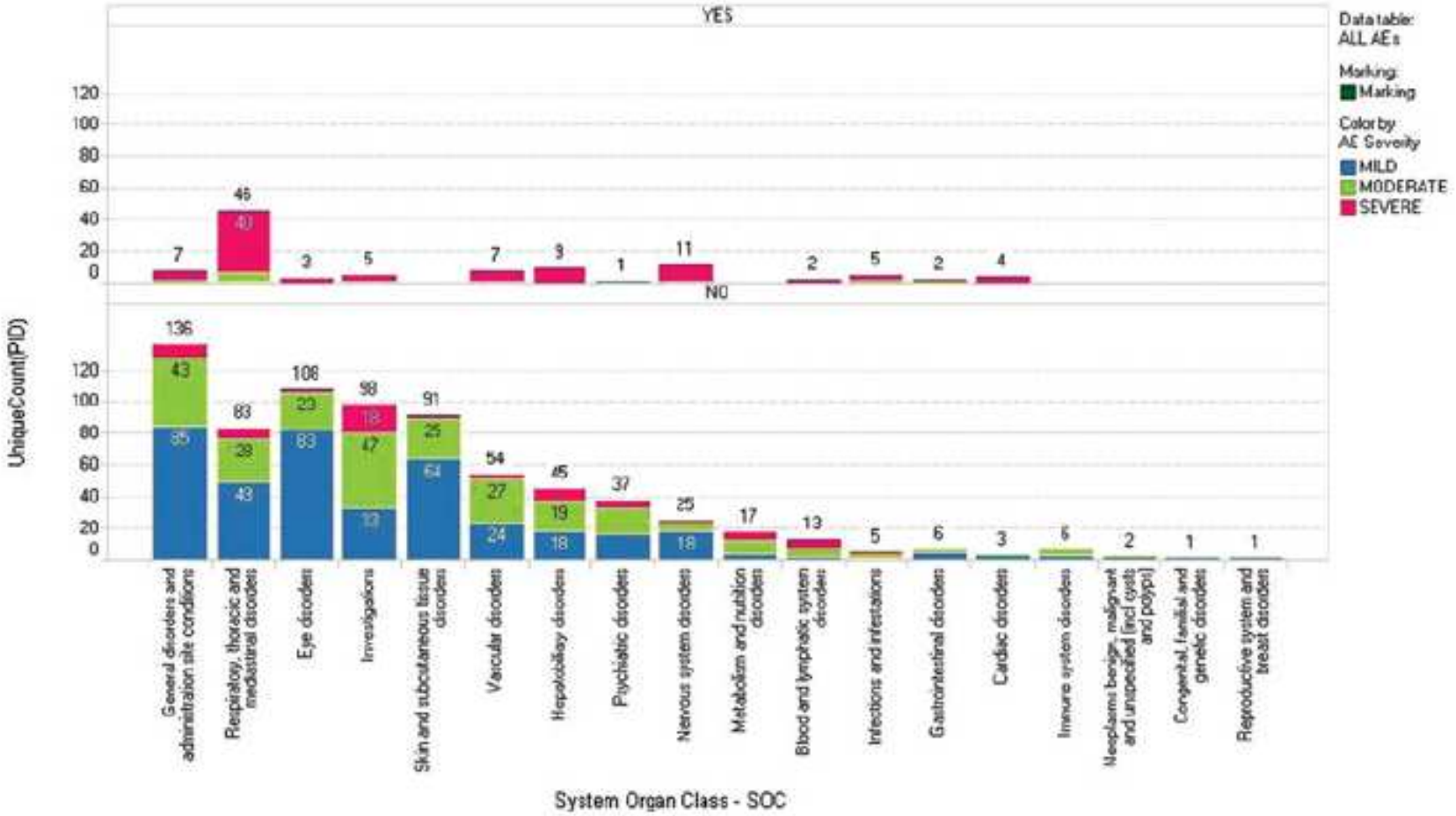
The data come into the form of:

- individual listings
- summary tables

"There is an urgent need to develop new methods that can summarize and deliver safety information in a format for more efficient and effective review and actions." (Chuang-Stein and Xia 2013) One proposal at Pfizer is based on the Spotfire software.

Example # 4: Spotfire software

From Chuang-Stein and Xia (2013):



The SPERT three-tier approach

Used at Pfizer, Amgen, and other companies. See Chuang-Stein and Xia 2013. AEs (Adverse Events) are classified into three tiers.

Tier-1 are those for which a pre-specified hypothesis has been specified. Formal statistical analysis specified in the protocol.

Tier-2 events are "common" but not pre-specified. Multiplicity adjustments often needed.

Tier-3 events are rare and are discussed case by case.

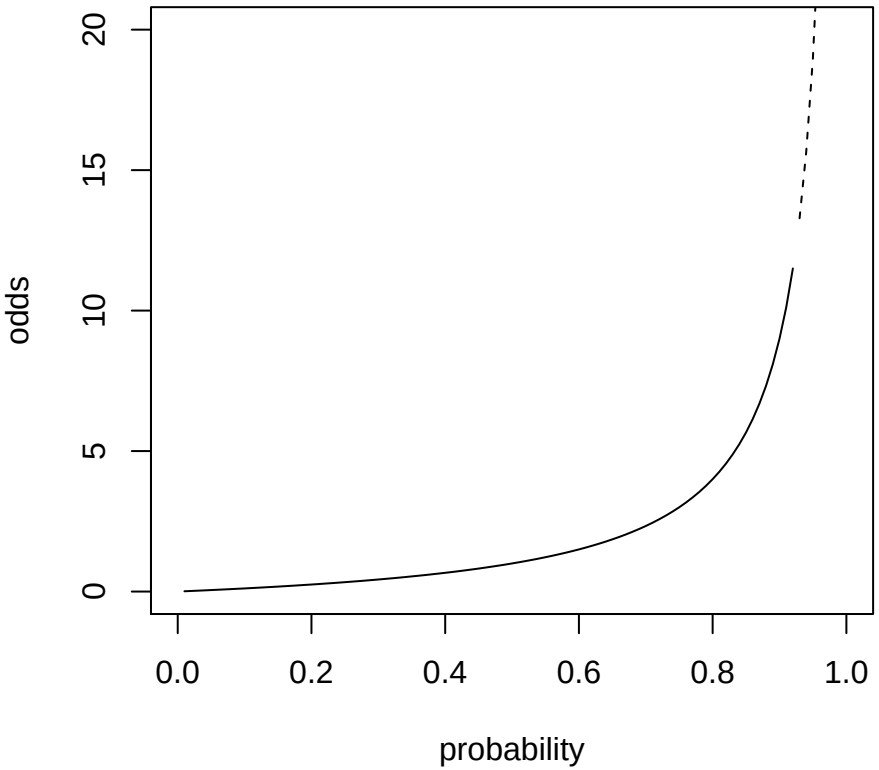
AEs are Binary endpoints

If p if the probability of the AE, the odds is the transformation

$$\text{odds}(p) = \frac{p}{1-p},$$

another way to express the probability of AE (used e.g. in betting). Its inverse is

$$p = \frac{\text{odds}(p)}{1+\text{odds}(p)}$$

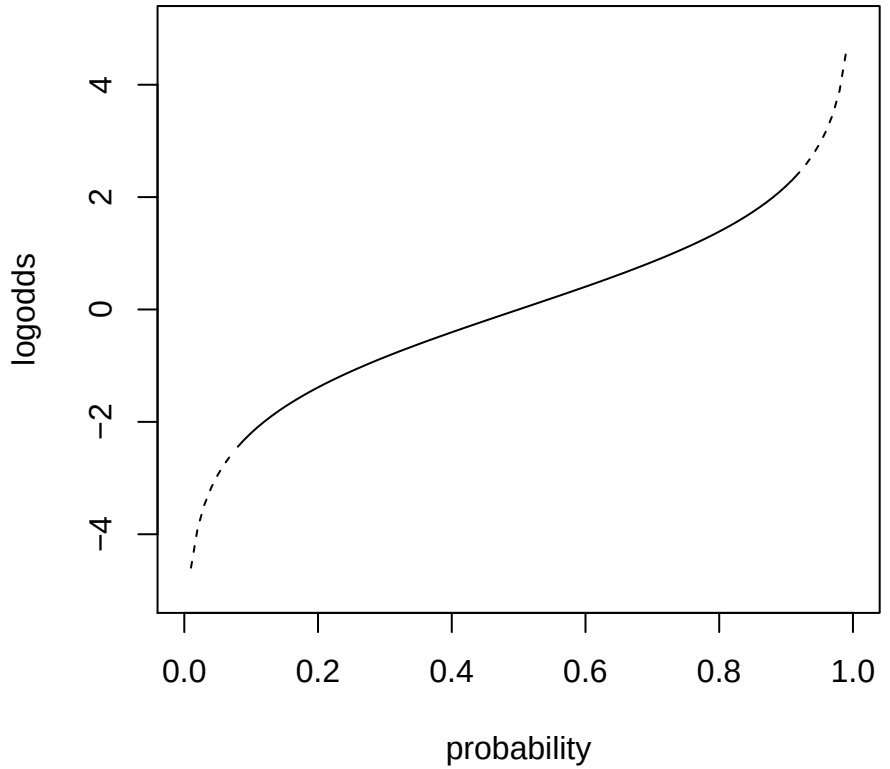


Probability and logit

Equivalently, the logit,
or logodds, transformation:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

is well defined for $0 < p < 1$.



Bayesian Idea #3: hierarchical priors to model multiplicity

In the MedDRA classification AEs are organized as preferred term (PT) according to System Organ Class (SOC). This provides a natural hierarchical structure that could be exploited in Bayesian Hierarchical Modelling.

AE rate estimates are pulled by shrinkage towards average values and at the same time they borrow strength from each other if they belogn to the same SOC.

The first to apply these Bayesian ideas to the analysis of binomial counts of AEs were Berry and Berry (2004): see next slide (but the corresponding DAG would have been better).

Extended by Xia *at al.* (2011) to Poisson likelihoods with varying times and exposures (and BUGS programs).

Example # 5: Berry and Berry model for binary safety data

A three-stage model for AE (expressed as PT).

The Likelihood

Count of PT j in SOC b under control: $X_{bj} \sim \text{Binomial}(N_c, c_{bj})$
Count of PT j in SOC b under treatment: $Y_{bj} \sim \text{Binomial}(N_t, t_{bj})$

The Hierarchical Prior

Stage 1 prior: logits of control rates (conditionally i.i.d. over j)

$$\text{logit}(c_{bj}) \sim \text{Normal}(\mu_{\lambda b}, \sigma^2_{\lambda b})$$

Stage 1 prior: logit differences of treatment rates (condit. i.i.d. over j):

$$\text{logit}(c_{bj}) - \text{logit}(t_{bj}) \sim p_b \ \delta(0) + \text{Normal}(\mu_{\theta b}, \sigma^2_{\theta b})$$

Stage 2 prior: independent normal-inverse gammas on $\mu_{\lambda b}, \sigma^2_{\lambda b}$ (conditionally i.i.d. over b):
Stage 2 prior: independent betas on p_b and independent normal-inverse gammas on $\mu_{\theta b}, \sigma^2_{\theta b}$ (conditionally i.i.d. over b):
Stage 3 priors: standard flat priors on the parameters of Stage 2

Bayesian Idea # 4: utility functions for Decision Making

The DMC activities and recommendations are only an ingredient in a series of Decisions in drug development, some of which can be taken based on quantitative arguments.

In Benefit-Risk (Efficacy/Safety) multicriteria decision analysis, for example, either deterministic or stochastic, several efficacy and safety measures are weighted to obtain a single evalutation. Classical decision-oriented utility functions may be useful.

Example # 6: stochastic multi-criteria decision analysis

Work in progress by Gaëlle Saint-Hilary

2 treatments: Treatment A and Treatment B, 1 study, 7 criteria

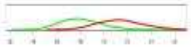
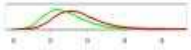
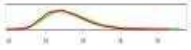
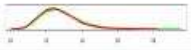
Criteria	Weights	Distributions of ξ_{ij} — Treatment A — Treatment B
Response	1/2	
Nausea	1/12	
Headache	1/12	
Diziness	1/12	
Insomnia	1/12	
Dry mouth	1/12	
Diarrhea	1/12	
$u(\xi_i, w)$ = **Benefit-Risk measure**		

Table : Criteria, Weights and B/R assessment

$\xi_{ij} \sim Beta(a, b)$, with $a = \#$ events, $b = \#$ patients - $\#$ events

Summary

Bayesian Idea # 1: use a community of priors

Bayesian Idea # 2: separate demands and beliefs

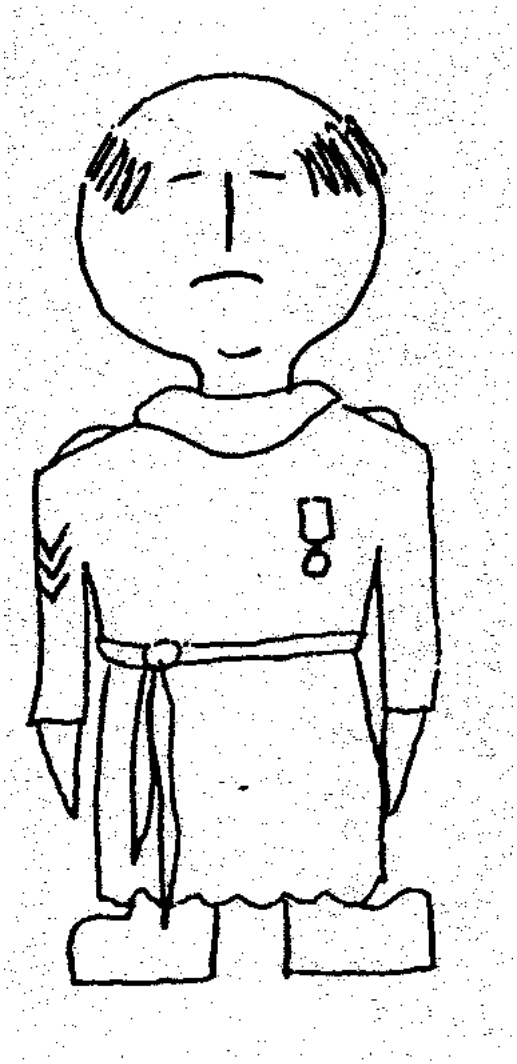
Bayesian Idea # 3: hierarchical priors to model multiplicity

Bayesian Idea # 4: utility functions for Decision Making

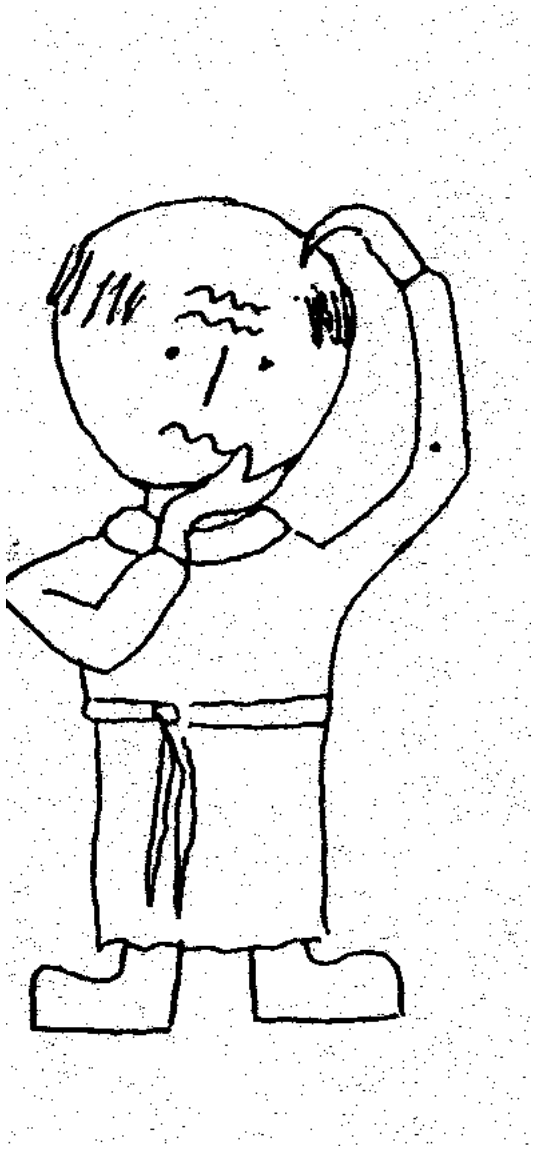
... plus six Examples given.

Courtesy of D.M. Titterington (1982)

Uniform Prior

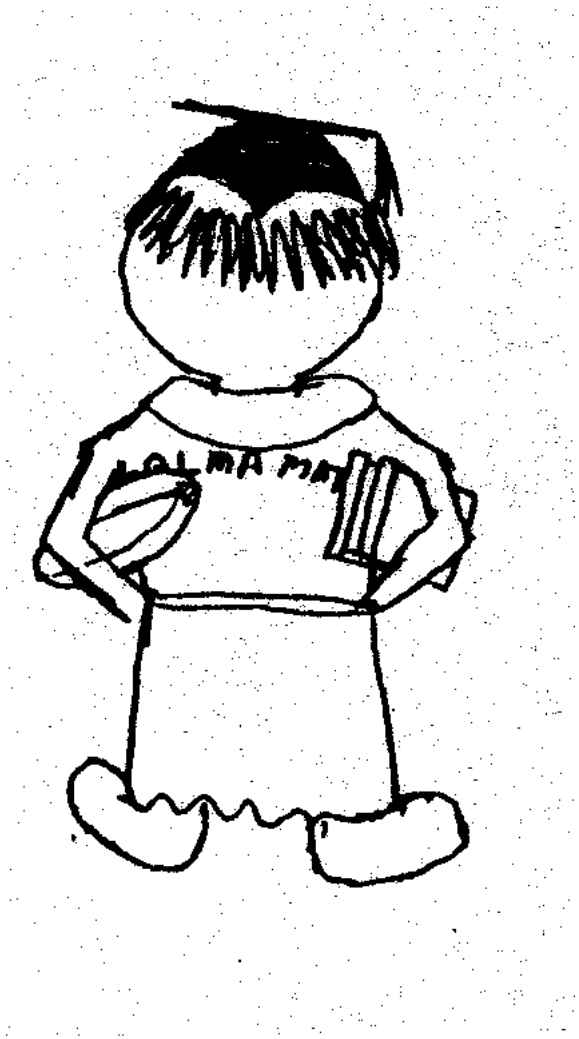


Vague Prior

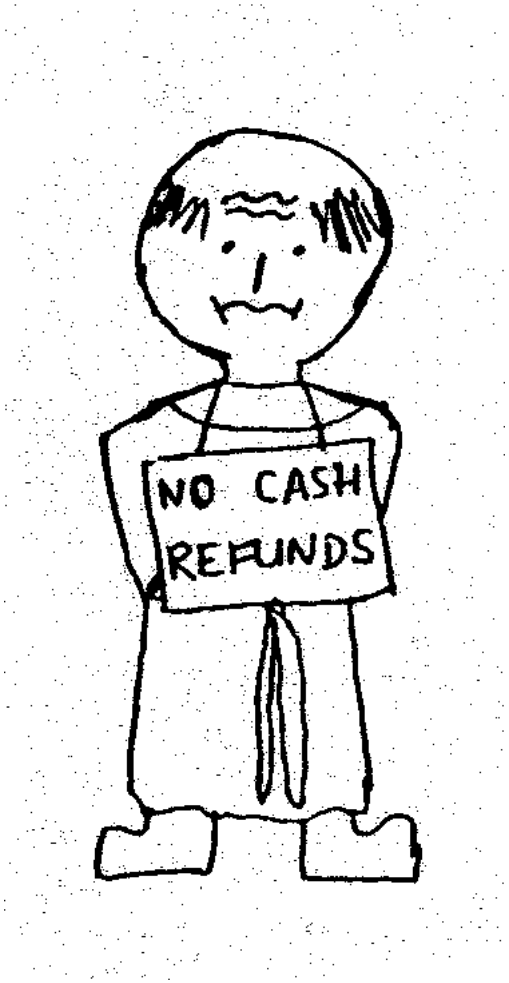


Informative Prior

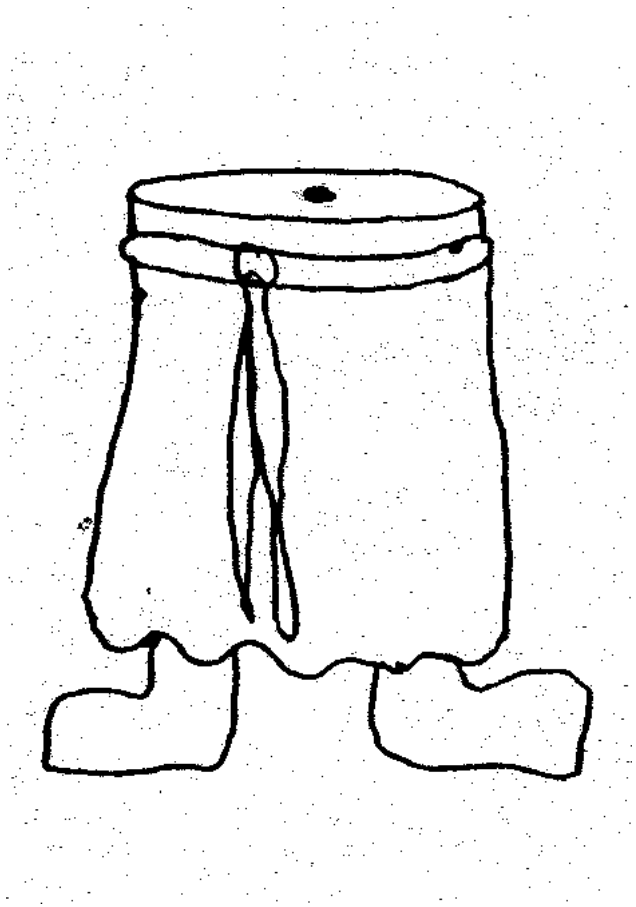
Student Prior



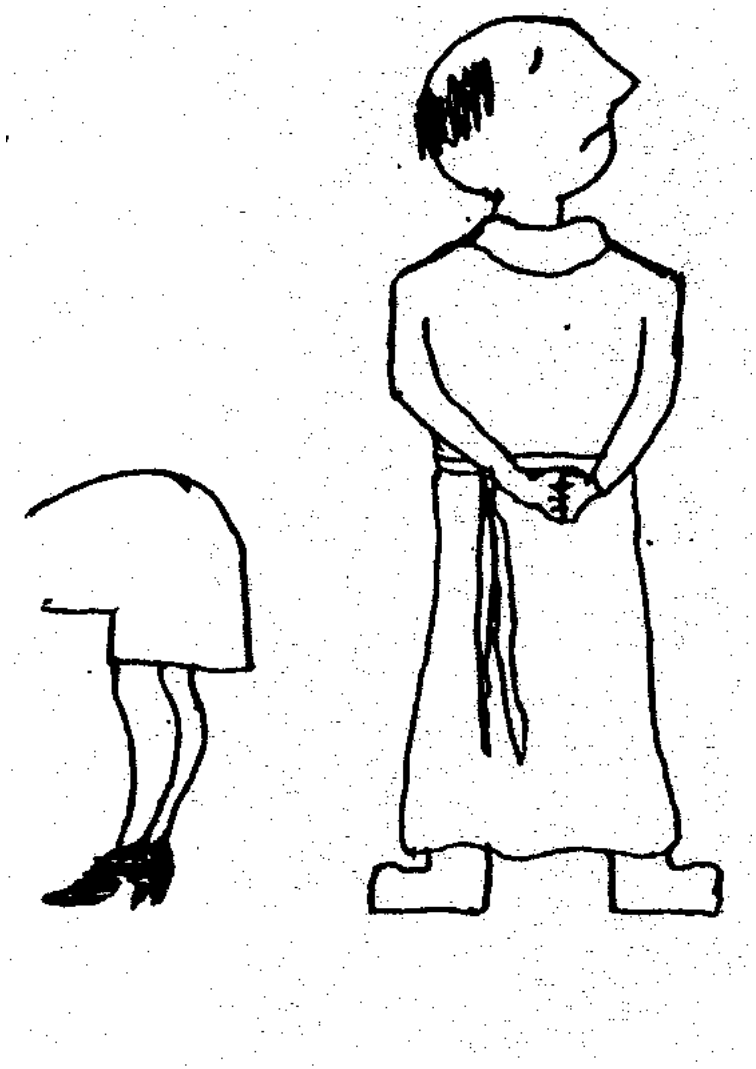
Exchangeable Prior



Truncated Prior

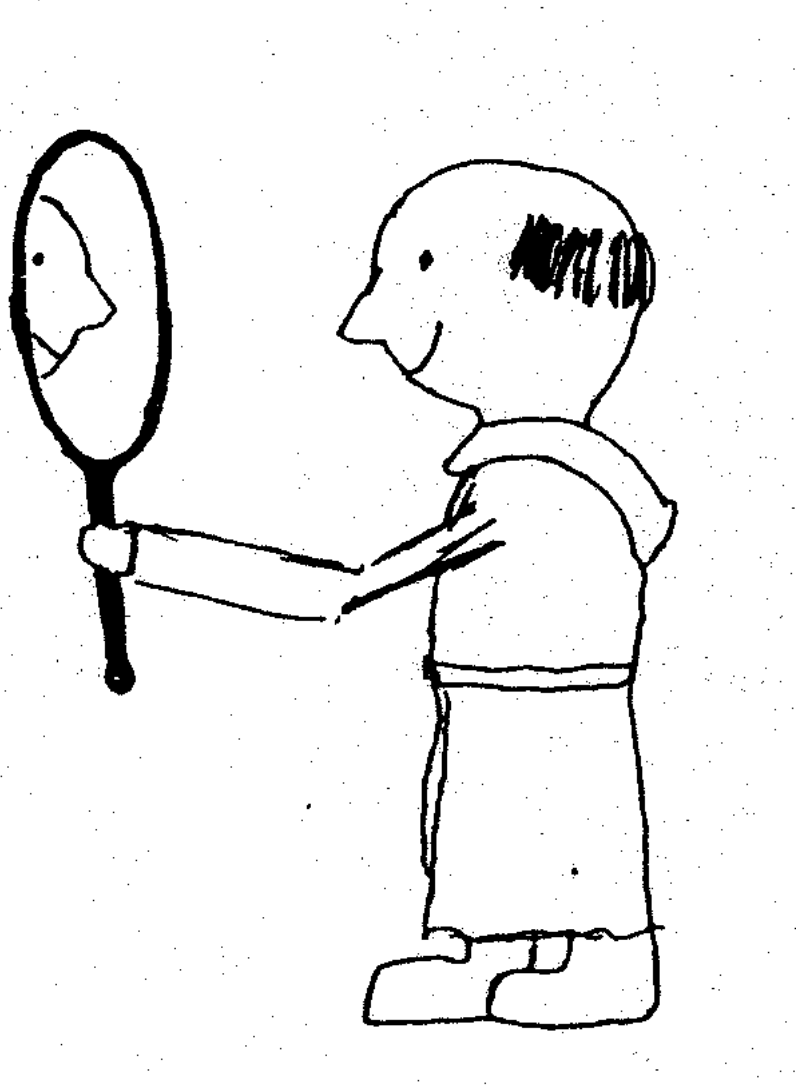


Proper Prior

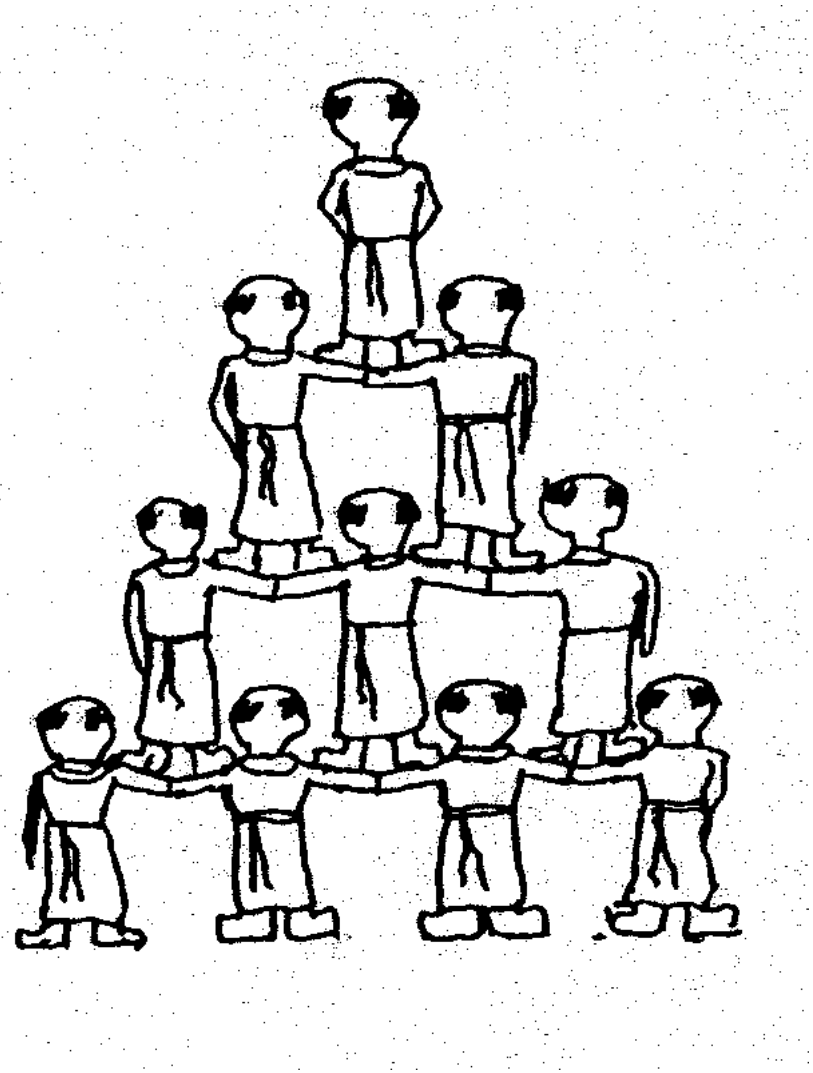


Improper Prior

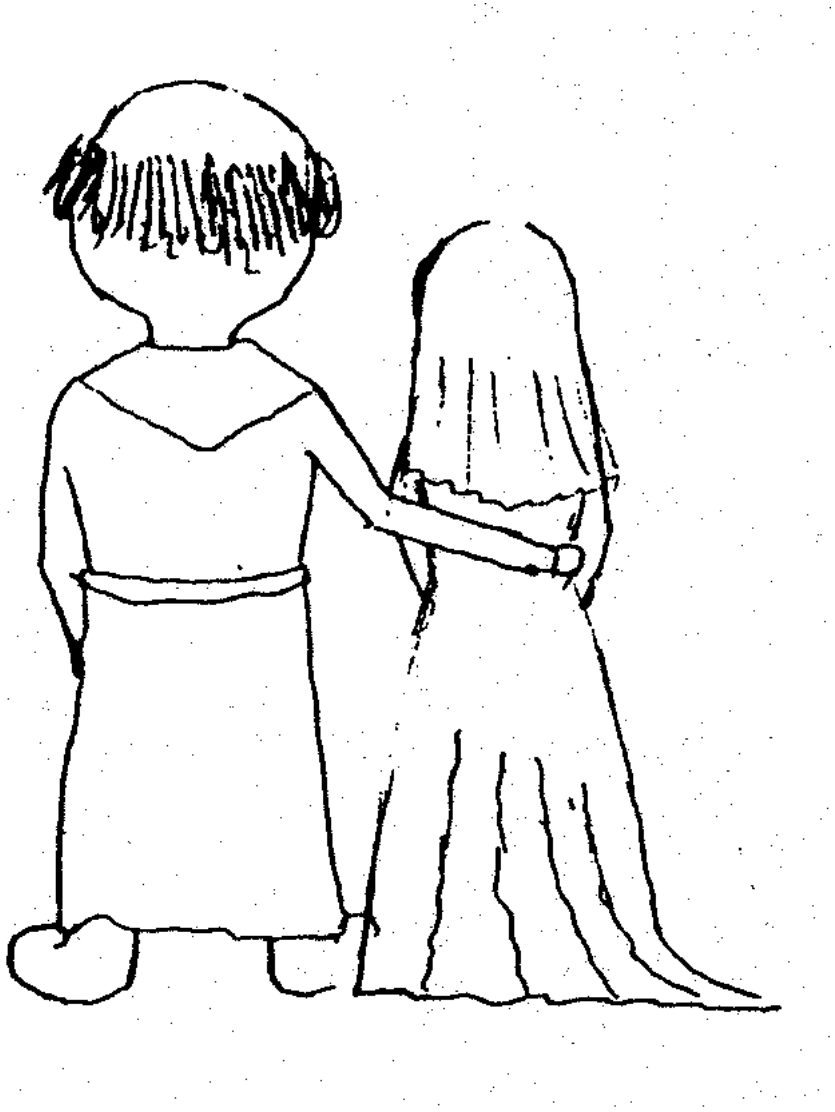
Subjective Prior



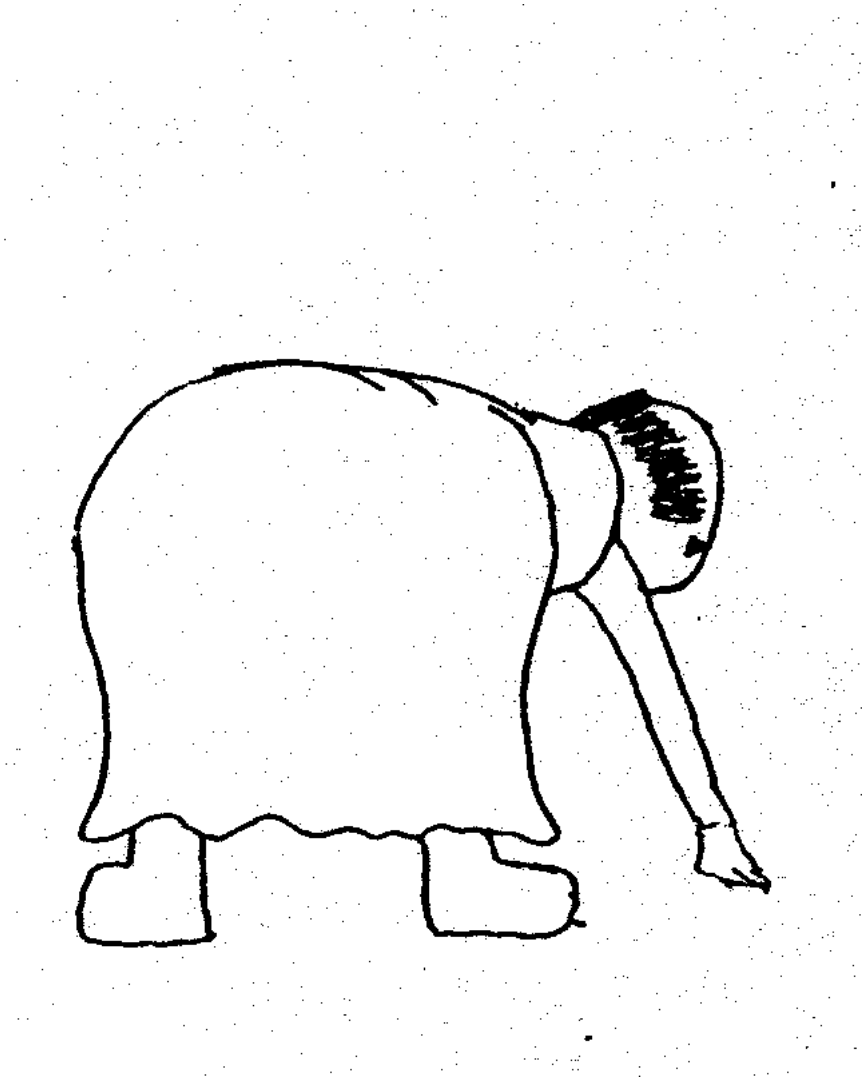
Hierarchical Prior



Conjugate Prior



... Posterior!!!



References

Berry SM Berry DA (2004). Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model. *Biometrics*, **60**, 418–426.

Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M and Racine-Poon A (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology *Statistics in Medicine* **28**, 1445–1463

Chuang-Stein C and Xia HA (2013) The Practice of Pre-Marketing Safety Assessment in Drug Development. *Journal of Biopharmaceutical Statistics* 23:1, 3-25.

Jennison C and Turnbull BW (2000). *Group sequential methods with applications in clinical trials*. Chapman & Hall.

Lunn DJ, Thomas A, Best N and Spiegelhalter D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and exten sibility. *Statistics and Computing* **10**, 325–337. `http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

Neuenschwander B, Rouyrre N, Hollaender N, Zuber E and Branson M. (2011) A proof of concept phase II non-inferiority criterion *Statistics in Medicine* **30**, 1618–1627.

Spiegelhalter D, Abrams KR, Myles J.P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Wiley.

Xia HA, Ma H and Carlin BP (2011). Bayesian Hierarchical Modeling for Detecting Safety Signals in Clinical Trials. *Journal of Biopharmaceutical Statistics*, 21:5, 1006-1029

Some more advertising

5th Early Phase Adaptive Trials Workshop

September 29 and 30, 2016 in Politecnico di Torino

Confirmed speakers:

Ying Kuen Cheung (Columbia University, USA)
Mark Conaway (University of Virginia, USA)
Elizabeth Garrett-Mayer (Medical University of South Carolina, USA)
Shing Lee (Columbia Universiy, USA)
Yimei Li (University of Pennsylvania, USA)
Thomas Jaki (Lancaster University, UK)
Adrian Mander (University of Cambridge, UK)
Xavier Paoletti (Gustave Roussy Cancer Campus, FR)
Mattew Schipper (University of Michigan, USA)
Peter Thall (M.D. Anderson Cancer Center, USA)
Mourad Tighiouart (Samuel Oschin Comprehensive Cancer Institute, USA)
Moreno Ursino (INSERM, FR)
Graham Wheeler (University of Cambridge, UK)
Christina Yap (University of Birmingham, UK)
Sarah Zohar (INSERM, FR)
Ziad Taib (Astrazeneca, Sweden)