Bayesian methods for the design of clinical trials in very rare diseases: application to the MYPAN trial in childhood polyarteritis nodosa

Lisa Hampson and John Whitehead
Lancaster University

Despina Eleftheriou, Paul Brogan
UCL Institute of Child Health

Joint DSBS/FMS Meeting,
2nd October 2014

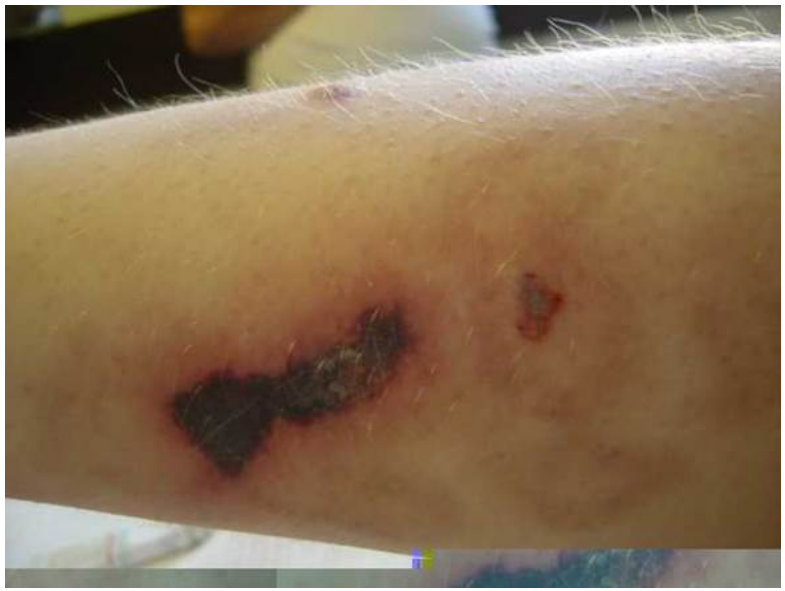
MRC Hubs for Trials Methodology Research
North West Hub

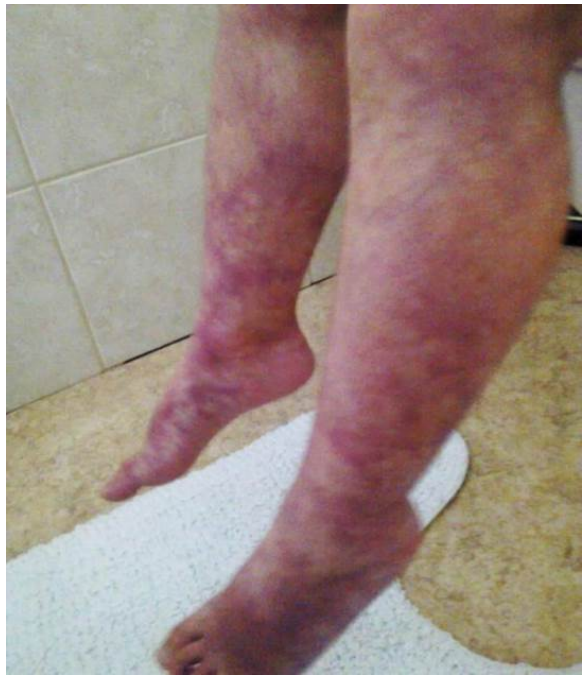
Medical and Pharmaceutical **Statistics Research Unit**

Outline

The MYPAN trial

Childhood polyarteritis nodosa (PAN) is a serious inflammatory blood vessel disease which affects around 1 per million children.





The MYPAN trial

- Treatment with cyclophosphamide (CYC) has been standard for the past 35 years. CYC is effective but toxic.
- Mycophenolate mofetil (MMF) is a new immunosuppresant which is thought to have a lower risk of toxicity.
- The MYPAN trial is an open-label RCT comparing MMF versus CYC for the treatment of PAN in children.
- The primary endpoint is remission within 6-months. Probabilities of remission on MMF and CYC are p_E and p_C. MMF will be preferred to CYC if $p_E - p_C \geq -0.1$.
- A definitive trial would require 513 patients per arm to have 90% power to declare MMF non-inferior to CYC when remission rates on both treatments equal 70%.

The MYPAN trial

- Treatment with cyclophosphamide (CYC) has been standard for the past 35 years. CYC is effective but toxic.
- Mycophenolate mofetil (MMF) is a new immunosuppresant which is thought to have a lower risk of toxicity.
- The MYPAN trial is an open-label RCT comparing MMF versus CYC for the treatment of PAN in children.
- The primary endpoint is remission within 6-months. Probabilities of remission on MMF and CYC are p_E and p_C. MMF will be preferred to CYC if $p_E - p_C \geq -0.1$.
- A definitive trial would require 513 patients per arm to have 90% power to declare MMF non-inferior to CYC when remission rates on both treatments equal 70%.

PROBLEM: 20-30 European centres could recruit 40 patients over 4 years.

What to do?

- If over a thousand patients are needed to reach a definitive conclusion, that is what is needed.
- Group sequential monitoring can achieve reductions in *expected* sample size of up to 40%, but this is not enough for the MYPAN trial
- We could settle for a less ambitious objective, which would be to improve our understanding of treatment options for PAN.

What to do?

- If over a thousand patients are needed to reach a definitive conclusion, that is what is needed.
- Group sequential monitoring can achieve reductions in *expected* sample size of up to 40%, but this is not enough for the MYPAN trial
- We could settle for a less ambitious objective, which would be to improve our understanding of treatment options for PAN.

WARNING: the following approach would not be used if there were sufficient patients for a conventional trial.

Bayesian approaches for clinical trials

Several authors have proposed Bayesian methods for rare disease trials: Lilford et al. (1995); Tan et al. (2003); Billingham et al. (2012).

Bayesian approaches for clinical trials

Several authors have proposed Bayesian methods for rare disease trials: Lilford et al. (1995); Tan et al. (2003); Billingham et al. (2012).

Informative prior distributions can be determined from historical data:

- Data on historical controls can be synthesised in a Bayesian random effects meta-analysis (Neuenschwander *et al*, 2010; Gsteiger *et al*, 2013).
- Historical data may be downweighted.
- Regulators are cautious about using informative prior distributions (EMEA, 2006).

Bayesian approaches for clinical trials

Several authors have proposed Bayesian methods for rare disease trials: Lilford et al. (1995); Tan et al. (2003); Billingham et al. (2012).

Informative prior distributions can be determined from historical data:

- Data on historical controls can be synthesised in a Bayesian random effects meta-analysis (Neuenschwander *et al*, 2010; Gsteiger *et al*, 2013).
- Historical data may be downweighted.
- Regulators are cautious about using informative prior distributions (EMEA, 2006).

No relevent published data were available to formulate a data-based prior for MYPAN.
Instead we elicited the beliefs of experts.

A Bayesian model
Hampson et al. (2014)

- We label MMF and CYC as treatments E and C, respectively.
- Primary endpoint is binary (success/failure). We represent the probability of success on E and C as p_E and p_C.

A Bayesian model
Hampson et al. (2014)

- We label MMF and CYC as treatments E and C, respectively.
- Primary endpoint is binary (success/failure). We represent the probability of success on E and C as p_E and p_C.

We measure the advantage of E over C using the log-odds ratio

$$\theta = \log\left\{\frac{p_E(1-p_c)}{p_C(1-p_E)}\right\}.$$

We prefer to work with the log-odds ratio, which is unconstrained, rather than the probability difference $p_E - p_C$, which must lie in the interval $[-1, 1]$.

A Bayesian model

The Bayesian approach begins by formally characterising prior opinion about p_C and θ.

Prior opinion about p_C and θ is assumed to be independent and modelled as:

- $p_C \sim \text{Beta}(a, b)$
- $\theta \sim N(\mu, \sigma^2)$.

The joint prior density of (p_C, p_E) can be found as

$$g_0(p_C, p_E) \propto \frac{p_C^{a-1}(1-p_C)^{b-1}}{p_E(1-p_E)} \exp\left(-\frac{1}{2\sigma^2}\left[\log\left\{\frac{p_E(1-p_C)}{p_C(1-p_E)}\right\} - \mu\right]^2\right).$$

A Bayesian model

The Bayesian approach begins by formally characterising prior opinion about p_C and θ.

Prior opinion about p_C and θ is assumed to be independent and modelled as:

- $p_C \sim \text{Beta}(a, b)$
- $\theta \sim N(\mu, \sigma^2)$.

The joint prior density of (p_C, p_E) can be found as

$$g_0(p_C, p_E) \propto \frac{p_C^{a-1}(1-p_C)^{b-1}}{p_E(1-p_E)} \exp\left(-\frac{1}{2\sigma^2}\left[\log\left\{\frac{p_E(1-p_C)}{p_C(1-p_E)}\right\} - \mu\right]^2\right).$$

Under the proposed model, prior opinion about p_C and p_E is correlated.

The marginal prior distribution of p_E does not follow a standard form but its density can be found from $g_0(p_C, p_E)$ using numerical integration.

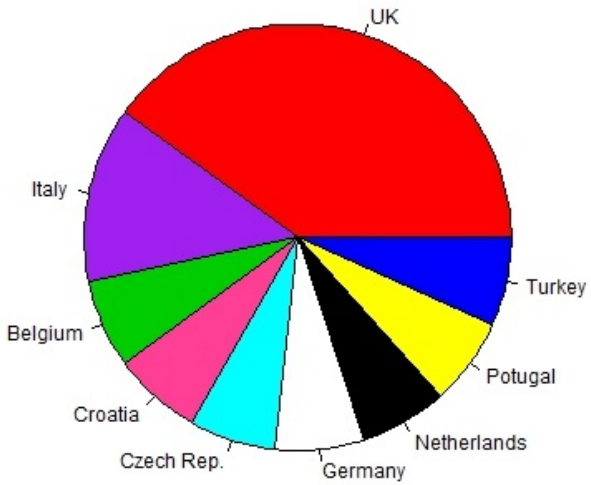
Identifying experts in childhood PAN

For the MYPAN trial, we defined an expert as a paediatric consultant

- Specialising in rheumatology, nephrology or immunology;
- With experience of treating children with PAN (on average 1 case every 2 years).

Experts were identified by sending invitations to society e-mail lists and paediatric clinics treating PAN identified via Orphanet (http://www.orpha.net)

15 experts from across the EU and Turkey attended 2-day prior elicitation meeting.



Structure of the Elicitation Meeting

Day 1 objectives:

- Provide experts with relevant training;
- Elicit expert opinion about p_C and θ.

Structure of the Elicitation Meeting

Day 1 objectives:

- Provide experts with relevant training;
- Elicit expert opinion about p_C and θ.

Day 2 objectives:

- Elicit expert opinion about relevance of unpublished RCT in related condition;
- Combine consensus opinion from Day 1 with related data;
- Check face validity of final consensus prior distributions.

Training the expert participants

The elicitation meeting began with training exercises including:

1. Clinical overview of PAN;
2. Overview of evidence supporting current treatments;
3. Introduction to Bayesian statistics.

We conducted a practice elicitation session, asking experts their opinion about the proportion of pink blocks in a jar which they were briefly shown.

This exercise was to intended to provide experts with experience of communicating uncertainty and interpreting prior distributions.

Day 1 formal elicitation exercise

We elicited experts' individual prior beliefs first before bringing the group together.

- Behavioural rather than mathematical aggregation of priors was preferred since different experts had different experiences and knowledge.

To elicit opinion about p_C and θ, experts were asked 6 questions.

Elicitation software

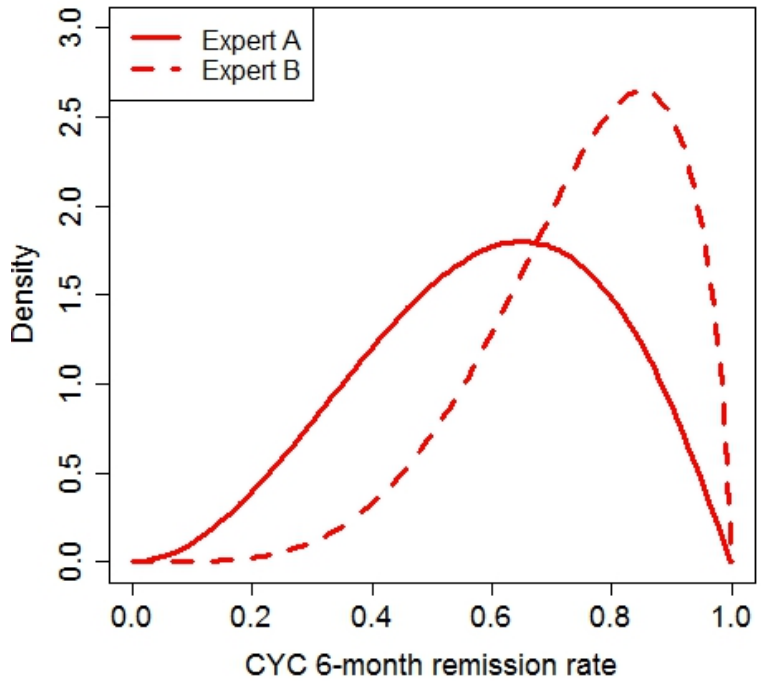
Once each expert had completed their questionnaire, they had a 1-to-1 meeting with a statistician who fed back:

1. plots of fitted probability density functions;
2. summaries of marginal priors (modes, means, credibility intervals);
3. strength of prior opinion (standard deviations; effective sample sizes (ESSs)).

Bespoke user-friendly software written in R using Shiny package.

Example opinion about p_C

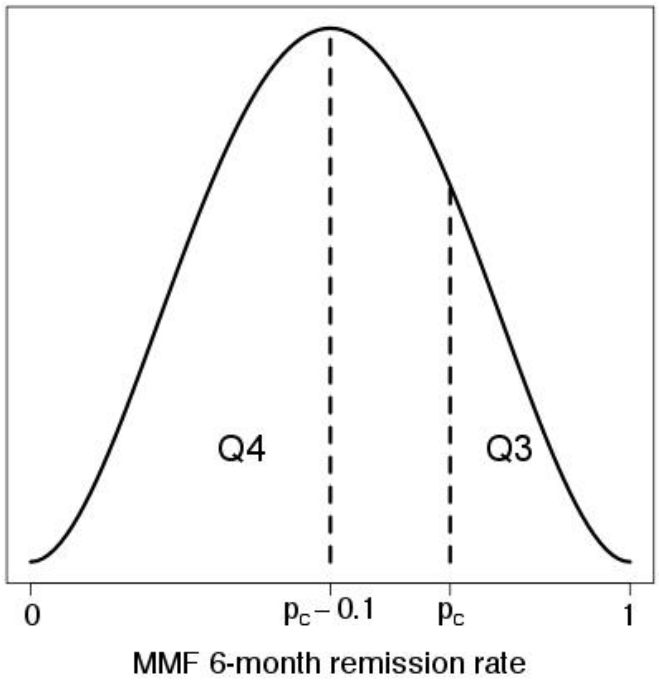
Q1: What do you think the 6-month remission rate for children with PAN on CYC is?
Q2: Provide a proportion such that you are 75% sure that the true 6-month remission rate on CYC exceeds this value.



Expert A: A1 = 0.65, A2 = 0.45; Expert B: A1 = 0.85, A2 = 0.65.

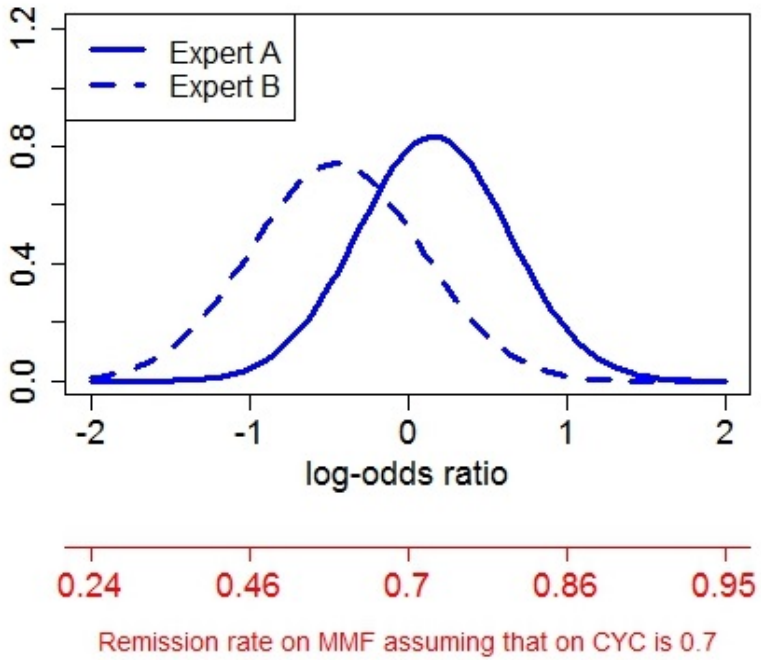
Example opinion about θ

Q3: What is chance that 6-month remission rate on MMF is higher than that on CYC?
Q4: What is chance that 6-month remission rate on CYC exceeds that on MMF by more than 10%?



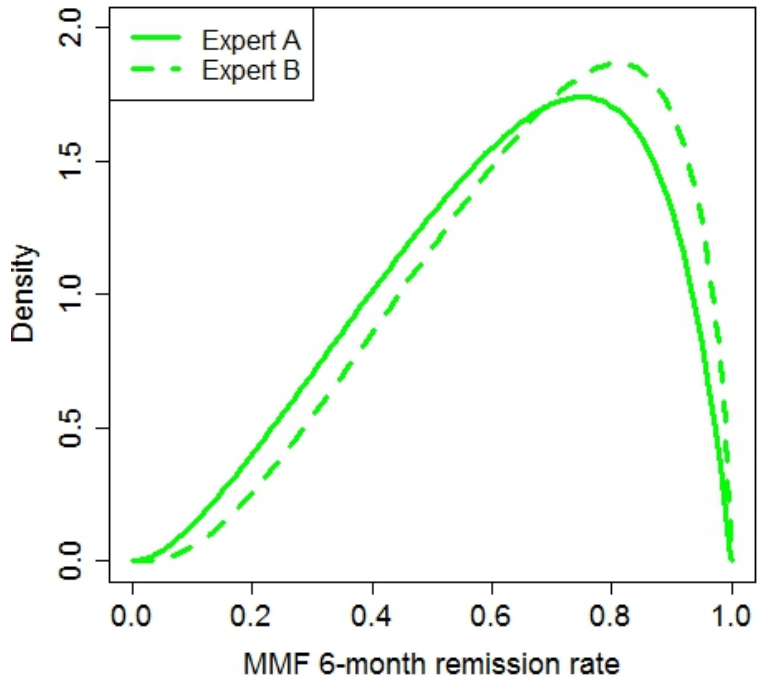
Example opinion about θ

Q3: What is chance that 6-month remission rate on MMF is higher than that on CYC?
Q4: What is chance that 6-month remission rate on CYC exceeds that on MMF by more than 10%?



Expert A: A1 = 0.63, A2 = 0.05; Expert B: A1 = 0.2, A2 = 0.4.

Example opinion about p_E



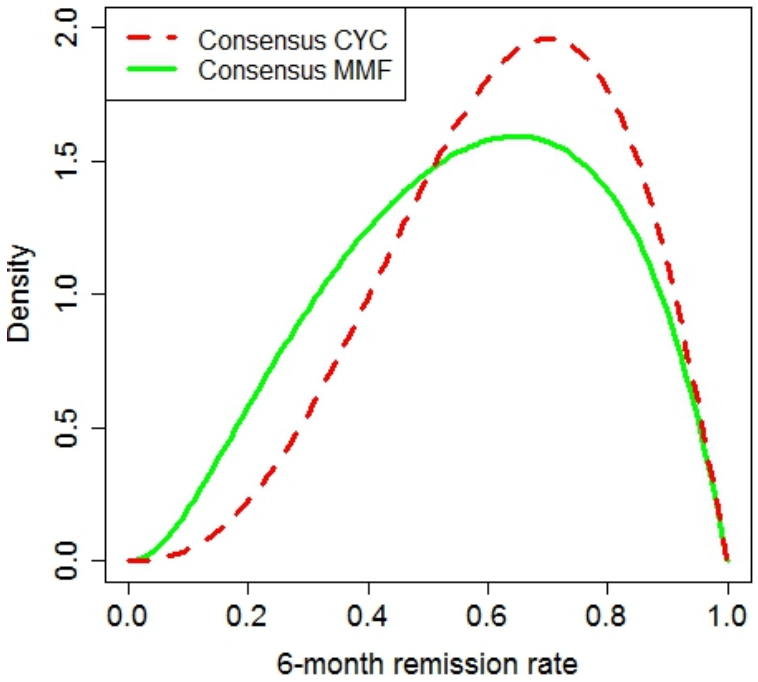
Properties of fitted p_E prior were compared with expert's initial answers to:

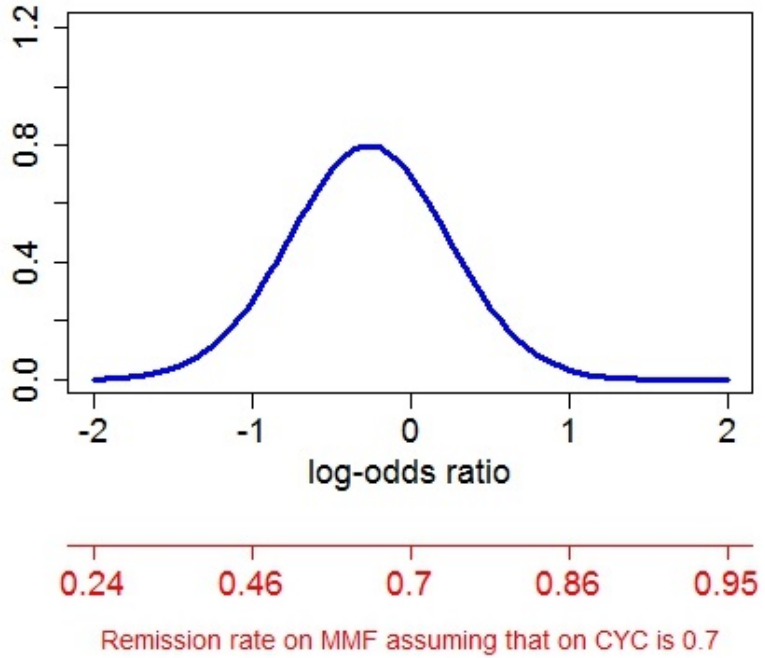
Q5: What do you think the 6-month remission rate for children with PAN on MMF is?
Q6: Provide a proportion such that you are 75% sure that the true 6-month remission rate on MMF exceeds this value.

Day 1 consensus prior distributions

Using a nominal group technique, we obtained consensus answers to Q1 - Q4. These answers specified prior distributions

$$p_C \sim \text{Beta}(3.6, 2.11) \quad \text{and} \quad \theta \sim N(-0.26, 0.25).$$





Consensus: A1 = 0.7, A2 = 0.5, A3 = 0.3, A4 = 0.3.

Effective sample sizes

Strength of prior opinion was characterised using ESSs (Morita et al., 2008).

These were influential in the group's final choice of consensus answers to Q1 - Q4.

Effective sample sizes

Strength of prior opinion was characterised using ESSs (Morita et al., 2008).

These were influential in the group's final choice of consensus answers to Q1 - Q4.

- p_C: ESS = size of single arm trial of CYC for which the expected Fisher's information for log-odds of success equals information represented by prior.
- θ: ESS = sample size needed for an RCT allocating equal numbers to MMF and CYC to have expected Fisher's information for θ equal to the information represented by stated prior.

Effective sample sizes

Strength of prior opinion was characterised using ESSs (Morita et al., 2008).

These were influential in the group's final choice of consensus answers to Q1 - Q4.

- p_C: ESS = size of single arm trial of CYC for which the expected Fisher's information for log-odds of success equals information represented by prior.
- θ: ESS = sample size needed for an RCT allocating equal numbers to MMF and CYC to have expected Fisher's information for θ equal to the information represented by stated prior.

Day 1 consensus prior opinion was equivalent to:

- 5 patients on CYC for p_C.
- 39 patients on each treatment for θ.

Combining opinion with related data

On Day 2, expert opinion was combined with data from the MYCYC trial, a soon-to-be published RCT comparing MMF versus CYC.

- MYCYC data were genuinely unknown to the experts on Day 1.
- MYCYC trial involved 132 adults and 8 children with a condition related to PAN.
- MYCYC primary endpoint was similar to MYPAN primary endpoint.

Combining opinion with related data

On Day 2, expert opinion was combined with data from the MYCYC trial, a soon-to-be published RCT comparing MMF versus CYC.

- MYCYC data were genuinely unknown to the experts on Day 1.
- MYCYC trial involved 132 adults and 8 children with a condition related to PAN.
- MYCYC primary endpoint was similar to MYPAN primary endpoint.

On Day 2, details of the MYCYC trial design were presented.

Before revealing the MYCYC results, experts were asked for their opinion about the relevance of these data for the MYPAN trial.

Combining expert opinion with related data

Individually, experts were asked:

1. What is the chance that the 6-month remission rate on CYC in the MYCYC patient group exceeds that in the MYPAN patient group?
2. What is the chance that the 6-month remission rate on CYC in the MYPAN patient group exceeds that in the MYCYC patient group by more than 10%?

Combining expert opinion with related data

Individually, experts were asked:

1. What is the chance that the 6-month remission rate on CYC in the MYCYC patient group exceeds that in the MYPAN patient group?
2. What is the chance that the 6-month remission rate on CYC in the MYPAN patient group exceeds that in the MYCYC patient group by more than 10%?

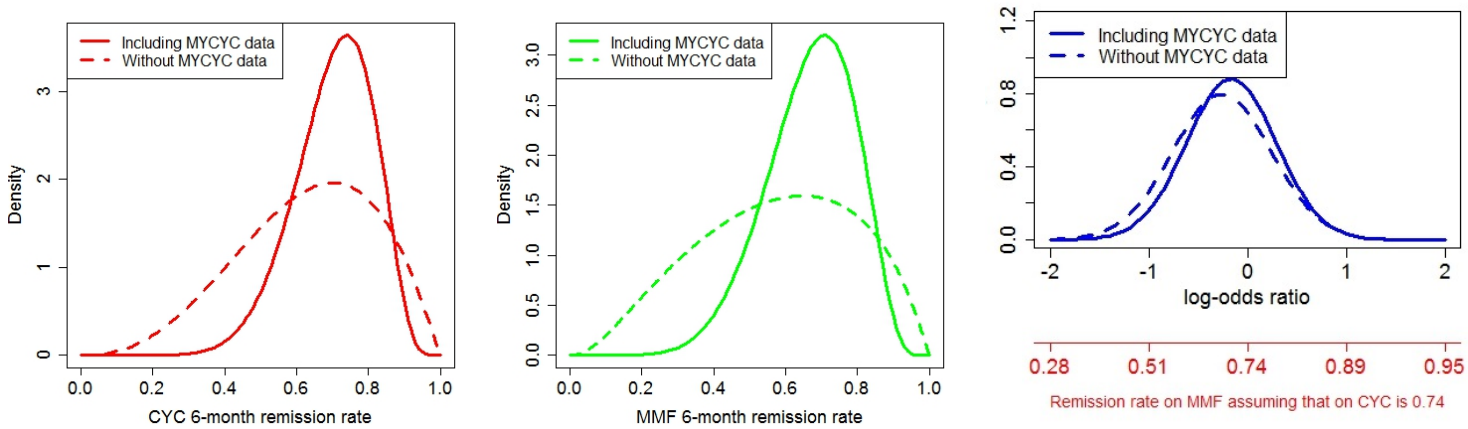
Two further questions, framed in terms of MMF, were asked.

Individuals' answers were then displayed on flip charts.

Consensus answers to 4 questions were proposed after each expert had explained their views.

Consensus priors incorporating MYCYC data

MYCYC results: 52/70 successes on CYC; 51/70 successes on MMF.



- Prior for p_C: mode = 0.74, 90% CI = (0.51, 0.86)
- Prior for p_E: mode = 0.71, 90% CI = (0.45, 0.85)
- Prior for θ: mode = -0.17, 90% CI = (-0.91, 0.58)
- Prior probability that MMF non-inferior to CYC is 0.77.

Discounting of the MYCYC data

Prior distributions for p_C and θ incorporating MYCYC data don't follow standard forms.

MYCYC data have a substantial influence on opinions about absolute values of success rates on the two treatments, but much less on their relative merits.

When prior densities incorporate consideration of the MYCYC data:

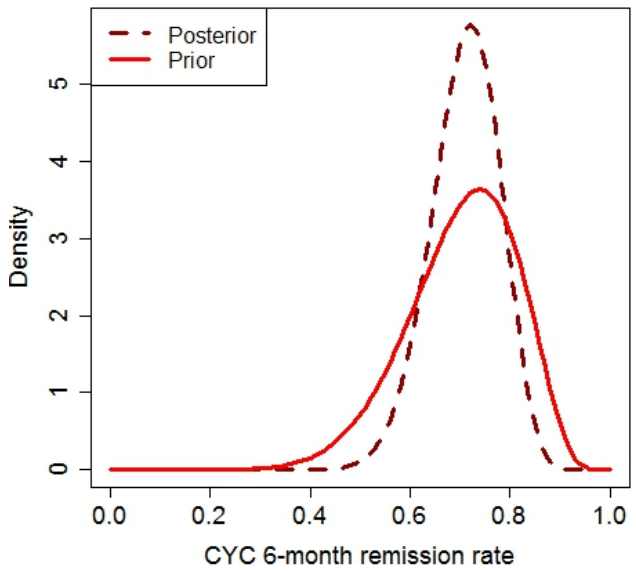
- ESS for p_C is 17 patients on CYC.
- ESS for θ is 48 patients on each treatment.

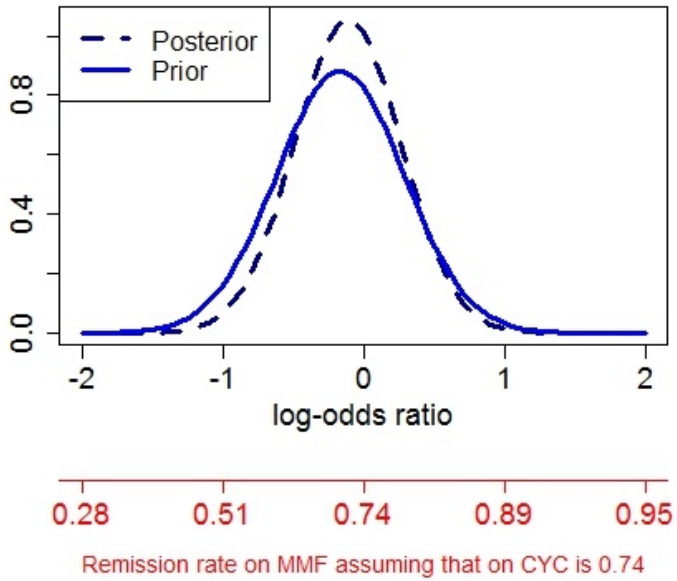
Judgements about the relevance of the MYCYC data mean that these data are discounted in the prior. 70 MYCYC patients per treatment increased the

- ESS for p_C by 12;
- ESS for θ by 9.

Combining opinion with hypothetical data

Suppose we observed $n_E = 20$, $S_E = 14$, $n_C = 20$, $S_C = 14$.





- Posterior for p_C: mode = 0.72, 90% CI = (0.59, 0.82)
- Posterior for p_E: mode = 0.70, 90% CI = (0.56, 0.80)
- Posterior probability that MMF non-inferior to CYC is 0.84.

Posterior decision rule

Calculate properties under decision rule which recommends E as non-inferior to C if

$$\Pi = \mathbb{P}\{p_E > p_C - 0.1 \mid \text{data}\} > 0.8.$$

Exact frequentist type I error rate

$$f(p_C, p_E) = \sum_{\{S_E, S_C; \Pi > 0.8\}} \binom{n_E}{S_E} \binom{n_C}{S_C} p_C^{S_C} (1 - p_C)^{n_C - S_C} p_E^{S_E} (1 - p_E)^{n_E - S_E}.$$

Exact Bayesian prior power

$$\frac{1}{\Pi_0} \int_0^1 \int_{\max\{0, p_C - 0.1\}}^1 f(p_C, p_E)\, g_0(p_E, p_C) \mathrm{d}p_E \mathrm{d}p_C.$$

For the MYPAN design with $n_E = n_C = 20$:

- Frequentist type I error rate is 0.29 under $p_E = 0.6$ and $p_C = 0.7$
- Bayesian prior power is 0.62.

Conclusions

- It is feasible to elicit prior opininon to inform the design of trials in rare diseases.
- For the MYPAN trial, experts accepted the Bayesian paradigm as a framework for representing their prior beliefs:
  - Most likely rates of disease remission on CYC and MMF are 74% and 71%.
  - Prior probability that MMF is non-inferior to CYC is 0.77.

Conclusions

- It is feasible to elicit prior opininon to inform the design of trials in rare diseases.
- For the MYPAN trial, experts accepted the Bayesian paradigm as a framework for representing their prior beliefs:
  - Most likely rates of disease remission on CYC and MMF are 74% and 71%.
  - Prior probability that MMF is non-inferior to CYC is 0.77.
- Experts are more confident about the relative merits of treatments but uncertain of absolute success rates.

Conclusions

- It is feasible to elicit prior opininon to inform the design of trials in rare diseases.
- For the MYPAN trial, experts accepted the Bayesian paradigm as a framework for representing their prior beliefs:
  - Most likely rates of disease remission on CYC and MMF are 74% and 71%.
  - Prior probability that MMF is non-inferior to CYC is 0.77.
- Experts are more confident about the relative merits of treatments but uncertain of absolute success rates.
- Consideration of how hypothetical data would shift prior opinion can allow us to demonstrate whether a small trial will influence opinion enough to change practice.

Conclusions

- It is feasible to elicit prior opininon to inform the design of trials in rare diseases.
- For the MYPAN trial, experts accepted the Bayesian paradigm as a framework for representing their prior beliefs:
  - Most likely rates of disease remission on CYC and MMF are 74% and 71%.
  - Prior probability that MMF is non-inferior to CYC is 0.77.
- Experts are more confident about the relative merits of treatments but uncertain of absolute success rates.
- Consideration of how hypothetical data would shift prior opinion can allow us to demonstrate whether a small trial will influence opinion enough to change practice.
- All possible outcomes of small trials can be enumerated. Thus we can evaluate the consequences of various designs (allocation ratios; decision rules).

References

Billingham L, Malottki K, Steven N. Small sample sizes in clinical trials: a statisticians perspective. *Clinical Investigation* 2012; **2**:655-7.

EMEA. Guideline on clinical trials in small populations. 2006.

Gsteiger S, Neuenschwander B, Mercier F, Schmidli H. Using historical control information for the design and analysis of clinical trials with overdispersed count data. *Statistics in Medicine* 2013; **32**:3609-22.

Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Statistics in Medicine*. In press.

Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. *British Medical Journal* 1995; **311**:1621-5.

Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. *Biometrics* 2008; **64**:595-602.

Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clinical Trials* 2010; **7**:5-18.

Tan S-B, Dear KBG, Bruzzi P, Machin D. Strategy for randomised clinical trials in rare cancers. *British Medical Journal* 2003; **327**:47-9.