# Bayesian methods to support clinical interpretation of treatment effects, interim decision making and handling of intercurrent events

Nicky Best (Biostatistics, GSK)
Prashant Dalvi (Clinical Research, GSK)

Views expressed are those of the authors and do not necessarily reflect those of EFSPI or GSK

EFSPI
EUROPEAN FEDERATION OF STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY
Representing Statistical Associations in Europe

# Case study – Ph2 study in chronic pain

- Phase 2, randomised, double-blind, placebo-controlled, dose-ranging trial investigating efficacy and safety of an IMP in **chronic pain**
  - ➢ Primary endpoint of weekly average of **daily pain scores using 11-point NRS** at week 12
  - ➢ Selection of endpoints/ population expected to be similar to Ph3 registrational studies

- Chronic pain trials are challenging, with a high and variable placebo response reported across trials

- Multiple examples of Ph3 trials with new MOA failing to meet the primary endpoint or showing efficacy of questionable clinical relevance

# Key clinical questions

1. Able to choose a dose that showed a **clinically relevant effect** vs placebo (not just statistically significant)
   - To optimize the PoS of demonstrating clinically compelling benefit in Ph3 on primary endpoint
   - Placebo-difference approximating MCID of ~1 units generally deemed convincing, but this may vary across different stakeholders
2. Able to stop the trial early for futility with an **interim analysis**
   - To ensure we avoid current and new patients being exposed to an ineffective treatment
   - Enable timely internal decision to stop the trial/ programme
3. Estimate the effect in patients where treatment discontinuations due to lack of efficacy/ AE and persistent use of prohibited therapies are considered a **negative outcome**
   - Composite strategy: post-ICE assessments to be imputed using multiple imputation based on baseline pain scores

What's the probability of the true treatment effect > clinically relevant difference?

Does it exceed a certain threshold?

# Trial design

- Ph2 dose finding, parallel group, placebo-controlled RCT
- Primary endpoint: change from baseline in average weekly pain score at week 12 (positive values = improvement)
- For simplicity, only consider 1 active dose v placebo here

# Issue 1: Clinically relevant treatment effect (CRT) – trial design

| | Conventional design |
|---|---|
| **Success rule for statistical significance** | P-value ≤ 0.025* for test of null hypothesis that true treatment effect ≤ 0 |
| **Type 1 error** | Controlled at 2.5% |
| **Minimum detectable effect#** | ~ 0.6 x CRT |
| **Success rule for clinical significance** | Option 1: Observed treat effect ≥ CRT<br><br>Option 2: P-value ≤ 0.025* for test of null hypothesis that true treat effect ≤ CRT |

* 1-sided

# If trial powered at 90% assuming true treatment effect = CRT

# Issue 1: Clinically relevant treatment effect (CRT) – trial design

| | Conventional design | Bayesian design (vague priors) |
|---|---|---|
| **Success rule for statistical significance** | P-value ≤ 0.025* for test of null hypothesis that true treatment effect ≤ 0 | Pr(true treatment effect > 0) ≥ 97.5% |
| **Type 1 error** | Controlled at 2.5% | Controlled at 2.5% |
| **Minimum detectable effect#** | ~ 0.6 x CRT | ~ 0.6 x CRT |
| **Success rule for clinical significance** | Option 1: Observed treat effect ≥ CRT<br><br>Option 2: P-value ≤ 0.025* for test of null hypothesis that true treat effect ≤ CRT | Pr(true treatment effect > CRT) ≥ 50%<br><br>Pr(true treatment effect > CRT) ≥ 97.5%<br><br>Pr(true treatment effect > CRT) ≥ 70% |

* 1-sided

# If trial powered at 90% assuming true treatment effect = CRT

# Issue 1: Clinically Relevant Treatment effect (CRT) – trial analysis

| | Conventional design | Bayesian design (vague priors) |
|---|---|---|
| **Primary analysis** | Frequentist MMRM | Bayesian MMRM with vague priors |
| **Estimated treatment effect:** difference vs placebo in mean improvement on pain score | Point estimate & 95% confidence interval<br>e.g. 0.88 (0.18, 1.57) | Posterior mean & 95% credible interval<br>e.g. 0.88 (0.18, 1.57)<br><br>Additional summaries<br>Pr(true treatment effect > 0) = 99.3%<br>Pr(true treatment effect > 0.5) = 85.8%<br>Pr(true treatment effect > 1) = 36.1%<br>Pr(true treatment effect > 1.2) = 17.7% |

# Issue 2: Interim analysis

- Interim futility analysis planned when ~50% participants complete week 12

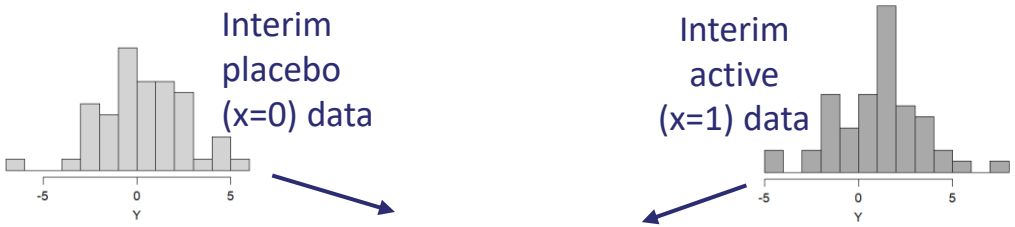
- Several methods for interim monitoring – metrics can be based on:
  - **Observed data** at interim (e.g. p-value of observing as or more extreme result under null; Bayesian posterior probability that treatment effect within meaningful interval)
  - Expected/**predicted future data** (e.g. conditional power; Bayesian predicted probability of success)

- Futility often framed as prediction problem – is there a low probability of trial meeting its objective?

- For pain study
  - Trial objective is to choose dose that demonstrates clinically relevant effect vs placebo
  - Interim is to assess if trial is unlikely to achieve this

# Conditional Power

**e.g. Stop if power (frequentist probability) to obtain statistical significance at EOS – given interim data and <u>assumed true treatment effect</u> (and SD) – is less that 20%**




$$\widehat{\mu(x)} = \overline{Y}_{interim}(x)$$

Interim Data / Future Data

| ID | $x$ | $y$ |
|---|---|---|
| 1 | 0 | -2 |
| … | | |
| K | 1 | 1 |
| K+1 | 0 | **0** |
| … | | |
| N-1 | 1 | **-1** |
| N | 1 | **-2** |

$\widehat{TE_1}, p_1 > 0.05$

*Fail*

| ID | $x$ | $y$ |
|---|---|---|
| 1 | 0 | -2 |
| … | | |
| K | 1 | 1 |
| K+1 | 0 | **1** |
| … | | |
| N-1 | 1 | **2** |
| N | 1 | **0** |

$\widehat{TE_m}, p_m > 0.05$

*Fail*

| ID | $x$ | $y$ |
|---|---|---|
| 1 | 0 | -2 |
| … | | |
| K | 1 | 1 |
| K+1 | 0 | **-2** |
| … | | |
| N-1 | 1 | **1** |
| N | 1 | **3** |

$\widehat{TE_M}, p_M$ <0.05

*Success*

*Success Proportion*:

Predicted probability of EOS statistical significance

# Conditional Power

**e.g. Stop if power (frequentist probability) to obtain statistical significance at EOS – given interim data and assumed true treatment effect (and SD) – is less that 20%**


$\widehat{\mu(x)} = \overline{Y}_{interim}(x)$

Interim Data / Future Data

| ID | x | y |
|---|---|---|
| 1 | 0 | -2 |
| ... | | |
| K | 1 | 1 |
| K+1 | 0 | 0 |
| ... | | |
| N-1 | 1 | -1 |
| N | 1 | -2 |

$\widehat{TE}_1, p_1 > 0.05$ → *Fail*

$\widehat{TE}_m, p_m > 0.05$ → *Fail*

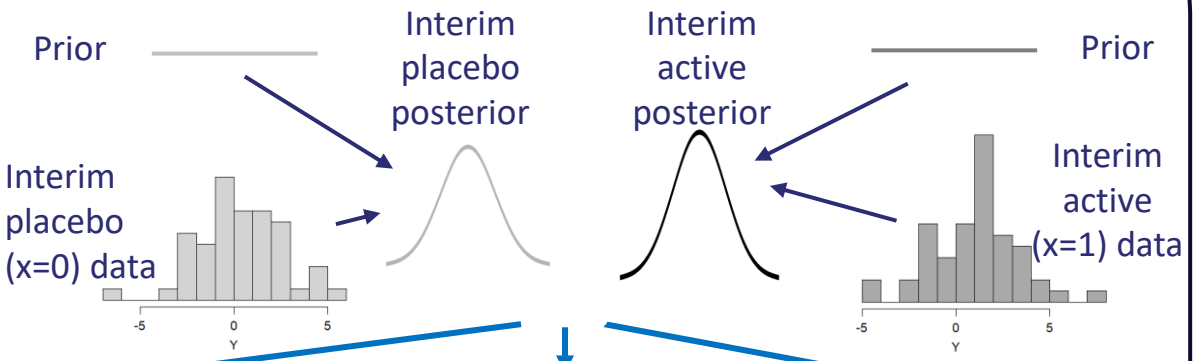
$\widehat{TE}_M, p_M < 0.05$ → *Success*

*Success Proportion:*

Predicted probability of EOS statistical significance

# Bayesian predicted PoS

**e.g. Stop if predicted probability – given interim data – that EOS posterior Pr(true effect > CRT) ≥ 70% is less than 15%**




$\mu(x)^{(1)}$ ... $\mu(x)^{(m)}$ ... $\mu(x)^{(M)}$ ...

| ID | x | y |
|---|---|---|
| 1 | 0 | -2 |
| ... | | |
| K | 1 | 1 |
| K+1 | 0 | -3 |
| ... | | |
| N-1 | 1 | 4 |
| N | 1 | 1 |

CRT >70% → *Success*

CRT <70% → *Fail*

CRT >70% → *Success*

*Success Proportion:*

Predicted probability of EOS success

Bayesian predicted power extends naturally to more complex prediction scenarios (e.g. inclusion of partial interim data, intercurrent event prediction)

# Issue 3: Intercurrent Event (ICE) strategy

- Treatment discontinuations due to lack of efficacy / AE and persistent use of prohibited therapies are considered a **negative outcome**
  - ➢ Composite strategy: post-ICE assessments to be imputed based on baseline pain scores
- Conventional analysis:
  1. Set post-ICE outcomes to missing
  2. Impute ‘bad outcomes’ via multiple imputation using baseline score distribution
  3. Fit frequentist MMRM model to imputed datasets
  4. Combine week 12 treatment effect estimates and SE using Rubin’s rules
- Theoretical justification for MI is as an approximation to Bayesian model
- Missing post-ICE outcomes handled automatically in Bayesian analysis by specifying appropriate prior (e.g. based on distribution of baseline scores)

# Summary

- Proposed fully Bayesian model can be designed to deliver:
  - ➢ pre-determined **frequentist operating characteristics** (e.g. 90% power with 5% false positive rate)
  - ➢ interim and EOS **decision rules** and **summaries of treatment benefit** that are closely **aligned with clinical reasoning**

  whilst automatically handling **analysis complexities** like ICE/missing data and partial interim data

**Would such a Bayesian design be acceptable in a pivotal Ph3 setting?**