# A Comparison of Some Methods for Detection of Safety Signals in Randomised Controlled Clinical Trials

Raymond Carragher

Project Supervisors:
Prof. Chris Robertson (University of Strathclyde)
Dr. Ian Bradbury (Frontier Science (Scotland))
Dr. David Young (University of Strathclyde)

The project is EPSRC funded with Frontier Science (Scotland) as the industrial partner.

# Overview

The main aim of this presentation is to:

- Compare (by way of a simulation) a number of existing approaches for analysing Adverse Events using groupings in clinical trials.
- Discuss the Adverse Event groupings and methods
- Look at a simulation study and the results
- Summary and conclusions

# Adverse Events

- Routinely recorded during a trial
- Severity - Common Terminology Criteria for Adverse Events provides a scale from 1–5 (1 = mild,..., 5 = death)
- Time of occurrence and/or duration
- Effect sizes may be small – long follow up / large numbers of patients
- Many different types of Adverse Events – may have multiple hypotheses

# Recent Approaches to Analysing Safety Data

A number of recent approaches to analysing safety data have grouped what they consider to be related adverse events into body-systems or System Organ Classes.

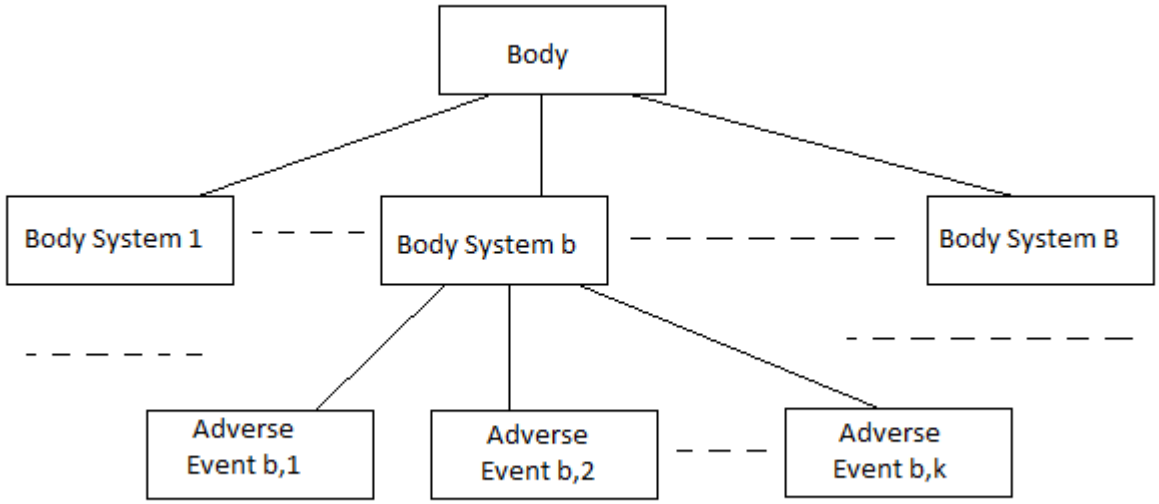
The idea being that if a treatment affects a particular body system then we may expect to see raised adverse event counts for all adverse events in that body-system.

- Berry, Berry - Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model (2004)

- Mehrotra, Adewale - Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals (2011).

# Body-System Hierarchy

The grouping by Body-System we consider has a natural hierarchical structure, with the body-system being part of an overall body and the adverse events being associated with particular body-systems:



# Methods

| Method | Description |
|---|---|
| HIER.BB | Berry and Berry model |
| HIER.1a | Subset of HIER.BB |
| BH | Control of the False Discovery Rate by the Benjamini-Hochberg procedure |
| DFDR | Double False Discovery Rate |
| NOADJ | Unadjusted testing |
| BONF | Bonferroni correction |
| GBH | Group Benjamini-Hochberg |
| ssBH | Subset Benjamini-Hochberg |

http://personal.strath.ac.uk/raymond.carragher/

Package: c212 – under development and untested

# Berry and Berry Model

***B*** body systems, body-system b containing $k_b$ Adverse Events

$N_C$ – number of patients in the control arm

$N_T$ – number of patients in the treatment arm

$X_{bj}$ – number of adverse events on the control arm

$Y_{bj}$ – number of adverse events on the treatment arm

AE counts: $X_{bj} \sim Bin(N_C, c_{bj})$, $Y_{bj} \sim Bin(N_T, t_{bj})$, $1 \le b \le B$, $1 \le j \le k_b$

Log Odds: $\gamma_{bj} = logit(c_{bj})$, $\theta_{bj} + \gamma_{bj} = logit(t_{bj})$

First Level: $\gamma_{bj} \sim N(\mu_{\gamma b}, \sigma^2_b)$, $\theta_{bj} \sim \pi_b I_{[0]} + (1 - \pi_b) N(\mu_{\theta b}, \sigma^2_b)$

$\theta_{bj}$ is the log odds-ratio for the occurrence of the AE for treatment compared to control. $\pi_b$ is the probability that there are no differences in rates in body-system *b.*

# Simulation Study

We used a simulation study to assess how the various methods performed with regard to detecting the raised levels of adverse events between treatment and control.

Model used to generate the simulated trial data:

$$logit(p_{tbj}) = \mu_{tbj} + U_{tbj}$$

$$X_{bj} \sim Bin(N_C, p_{1bj})$$

$$Y_{bj} \sim Bin(N_T, p_{2bj})$$

where $\mu$ is a fixed effect and $U$ is a random effect.

# Simulation Study

Results from one particular (repeated) simulation:

8 body-systems with between 1 and 11 adverse events in each body-system.
45 adverse events in total.

Trials size:
Trial 1 – 110 patients in each arm
Trial 2 – 450 patients in each arm
Trial 3 – 1100 patients in each arm

For all trials:
The AE rate was raised for body-system 5 for both treatment and control.
The AE rate was raised for body-system 3 for treatment only.
The AE rate was raised for body-system 2 for two out of 4 AEs for treatment only.

# Simulation Study

500 simulations in total.

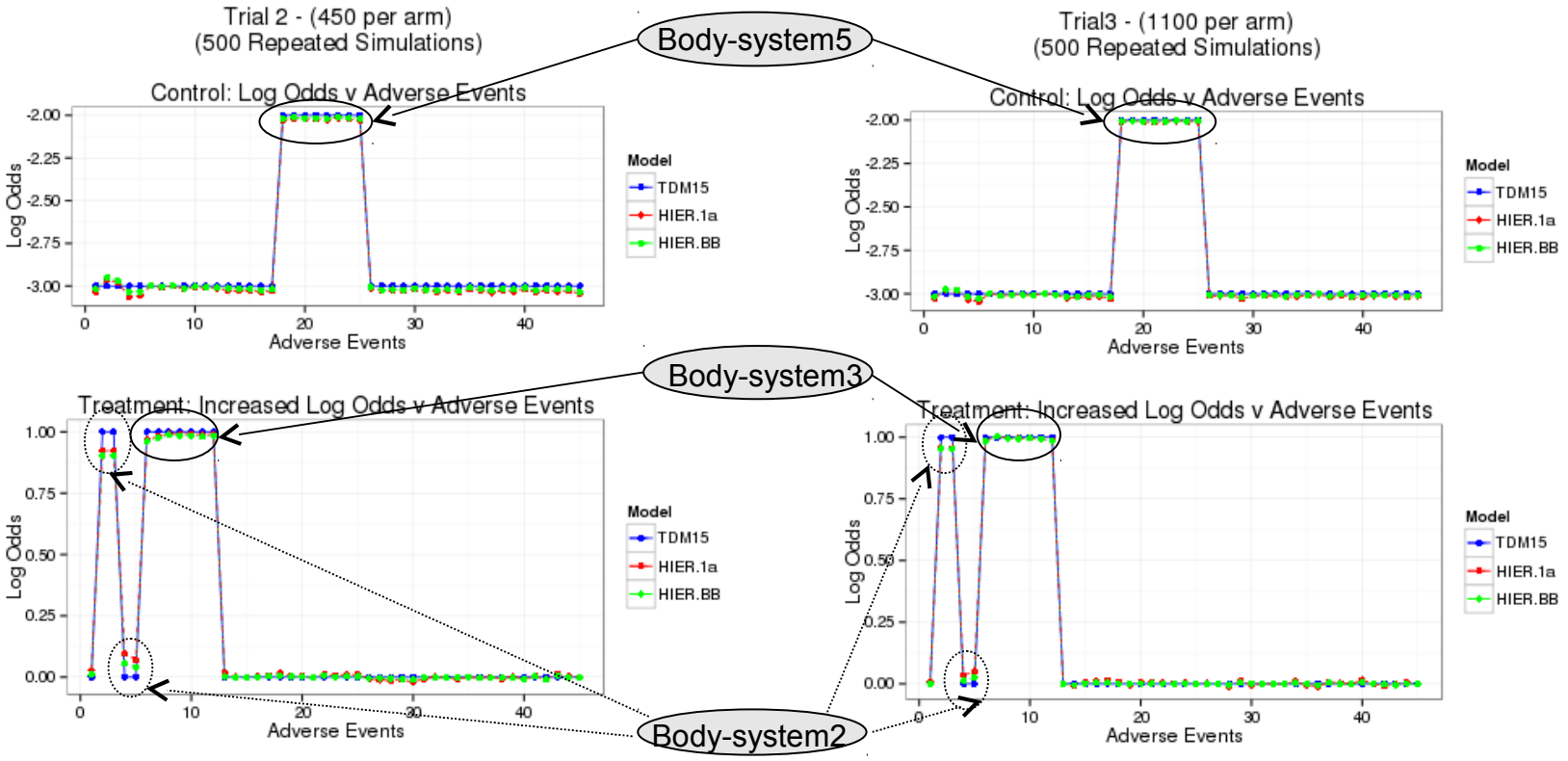
**Adverse Event Numbers**
In each simulation there are 9 Adverse Events which have underlying rate raised in treatment compared to control.

22500 Adverse Events over the whole simulation.

4500 Adverse Events with raised rates over the whole simulation.

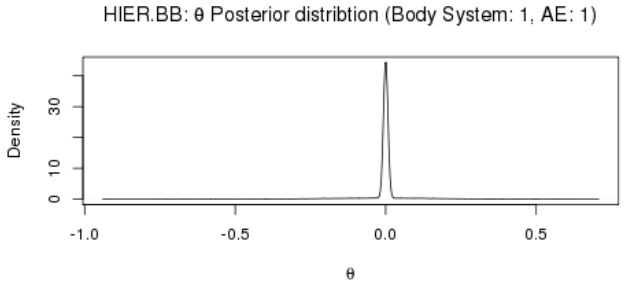
**Flagging an Adverse Event:**
95% posterior probability for Bayesian methods
5% significance level for the error controlling methods

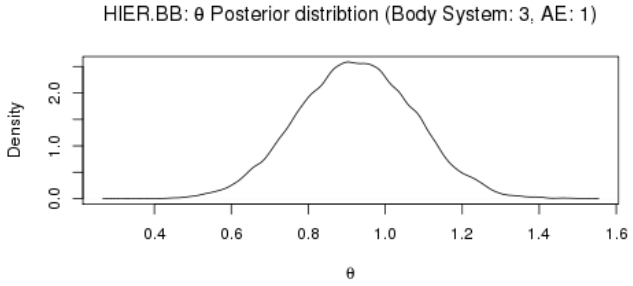
# Simulation Study



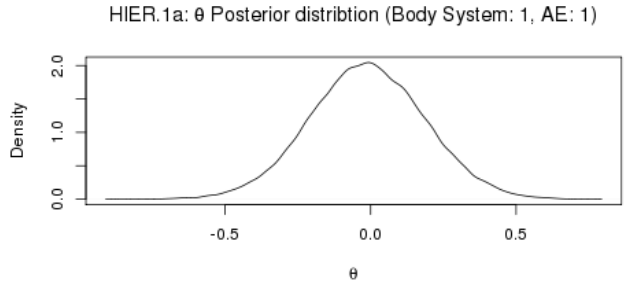
# Simulation Study

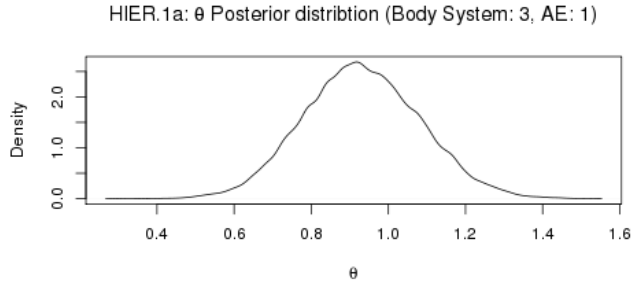
Berry & Berry Model:
HIER.BB (point mass):




HIER.1a (no point mass):



# Simulation Study – Trial 2 (450 per arm)

| Method | Correct | Incorrect | Missed |
|---|---|---|---|
| Berry & Berry (HIER.BB) | 4303 | 9 | 197 |
| Berry & Berry without point mass (HIER.1a) | 4492 | 582 | 8 |
| Unadjusted Testing (NOADJ) | 4374 | 682 | 126 |
| Bonferroni Correction (BONF) | 3258 | 12 | 1242 |
| Double False Discovery Rate (DFDR) | 4317 | 72 | 183 |
| False Discovery Rate (BH) | 4022 | 114 | 478 |
| Group Benjamini-Hochberg (GBH) | 4441 | 144 | 59 |
| Subset Benjamini-Hochberg | 3848 | 14 | 652 |

# Simulation Study – Trial 3 (1100 per arm)

| Method | Correct | Incorrect | Missed |
|---|---|---|---|
| Berry & Berry (HIER.BB) | 4498 | 5 | 2 |
| Berry & Berry without point mass (HIER.1a) | 4500 | 705 | 0 |
| Unadjusted Testing (NOADJ) | 4500 | 707 | 0 |
| Bonferroni Correction (BONF) | 4486 | 10 | 14 |
| Double False Discovery Rate (DFDR) | 4500 | 67 | 0 |
| False Discovery Rate (BH) | 4499 | 132 | 1 |
| Group Benjamini-Hochberg (GBH) | 4500 | 143 | 0 |
| Subset Benjamini-Hochberg | 4498 | 25 | 2 |

# Summary

- The simulations have indicated that where there are relationships between the Adverse Events using groupings (body-systems) do appear to make a difference to the results.

- The point mass in the Berry & Berry model (HIER.BB) makes a quantitative difference to the results.

- For the error controlling methods it may be difficult to objectively pick a method of analysing the data before the trial.

- The body-system described in Berry & Berry looks to be a worthwhile structure to consider when modelling data.

- The models and data discussed here do not take into account the severity of events.

- The models and data discussed here do not take into account the timings of events.

# References


[1] D. V. Mehrotra and J. F. Heyse. Use of the false discovery rate for evaluating clinical safety data. Stat Methods Med Res, 13(3):227–38, 2004.

[2] S. M. Berry and D. A. Berry, “Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model,” Biometrics, vol. 60, no. 2, pp. 418–426, 2004.

[3] D. V. Mehrotra and J. F. Heyse. Use of the false discovery rate for evaluating clinical safety data. Stat Methods Med Res, 13(3):227–38, 2004.

[4] D. B. Dunson, A. H. Herring, and S. M. Engel, “Bayesian selection and clustering of polymorphisms in functionally related genes,” Journal of the American Statistical Association, vol. 103, no. 482, pp. 534–546, 2008.

[5] J. N. S. Matthews, Introduction to Randomized Controlled Clinical Trials, Second Edition. Chapman & Hall/CRC Texts in Statistical Science, Chapman and Hall/CRC, 2006.

[6] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1):289–300, 1995.

[7] J. X. Hu, H. Zhao, and H. H. Zhou. False discovery rate control with groups. J Am
Stat Assoc, 105(491):1215–1227, 2010.

[8] Daniel Yekutieli. False discovery rate control for non-positively regression dependent test statistics. Journal of Statistical Planning and Inference, 138(2):405–415, 2008.

[9] H. Amy Xia, H. Ma, and B. P. Carlin, “Bayesian hierarchical modeling for detecting safety signals in clinical trials,” Journal of Biopharmaceutical Statistics, vol. 21, no. 5, pp. 1006–1029, 2011.