Detection of Safety Signals in Randomised Controlled Trials using Groupings

Raymond Carragher
University of Strathclyde
Supervisors: Prof. Chris Robertson, Dr. Ian Bradbury (Frontier Science Scotland Ltd.), Dr. David Young.

1. Introduction

Many different types of adverse event are routinely recorded during clinical trials. The statistical analysis of this data may need to take into account:

1.potential multiple comparison issues;
2.low power - effect sizes of adverse events in clinical trials are generally small.

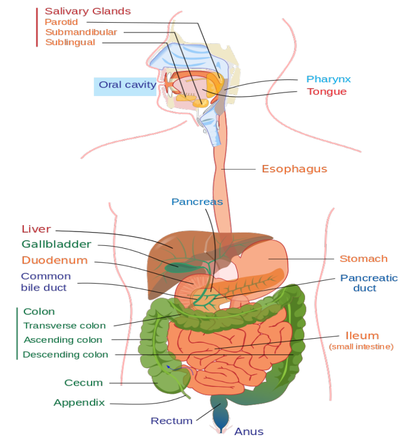
The use of methods which use possible groupings of adverse events (e.g. by System Organ Class) in their statistical analyses may result in an increase in the power to detect adverse event incidence while maintaining control over the Type-I error rate.

2. System Organ Class

Medical dictionaries (e.g. MedDRA) provide groupings of adverse events by System Organ Class (SOC).

Gastrointestinal Disorders

Diarrhoea
Nausea
Vomiting
Abdominal pain
Stomatitis
⋮
Haematemesis



3. Methods

Berry and Berry (2004): A Binomial Bayesian three-level hierarchical model where the increase in log-odds (θ_{bj}) of the occurrence of an adverse event under treatment is modelled as a mixture distribution:

$$X_{bj} \sim \text{Bin}(N_C, c_{bj})$$

$$Y_{bj} \sim \text{Bin}(N_T, t_{bj})$$

$$\theta_{bj} \sim \pi_b \, \text{I}_{[\theta_{bj}=0]} + (1-\pi_b) \, \text{I}_{[\theta_{bj}\neq 0]} \, \text{N}(\mu_{\theta b}, \sigma^2_{\theta b})$$

The system organ classes and averse events are indexed by *b* and *j* respectively, and each system organ class has a common mean and variance.

Double False Discovery Rate (DFDR) (2012): Application of the False Discovery Rate at both the System Organ Class and individual adverse event level.

Group Benjamini-Hochberg (GBH) (2010): A p-value weighted application of the False Discovery Rate where groupings of hypotheses are used to calculate the weightings.

4. Clinical Trial Safety Study

23 SOCs, 497 types of adverse event. Diarrhoea and Rash were expected adverse events based on Phase I/II studies. At the end of the trial 10 adverse events were significant at 5% level for a Fisher exact test comparing treatment with control.

Adverse Events	p-value	Berry and Berry model¶	DFDR¤	GBH¤
Diarrhoea	**<0.001***	**1.000**	**Y**	**Y**
Rash	**<0.001***	**1.000**	**Y**	**Y**
Epistaxis	0.004	**0.980**	N	N
Dyspepsia	0.004	**0.986**	N	N
Dermatitis acneiform	0.008	**0.967**	N	N
Muscle spasms	0.035	0.892	N	N
Localised infection	0.038	0.772	N	N
Arthralgia	0.039	0.905	N	N
Back pain	0.047	0.879	N	N
Nail disorder	0.049	0.941	N	N

* Remains significant after the application of the Bonferonni correction.
¶ Posterior probability that the change in log-odds of the occurrence of the adverse event on the treatment arm is positive.
¤ Flagged as significant by the procedure at the 5% or 10% significance level (Y = yes, N = no).

Under the Berry and Berry model 5 adverse events have posterior probability exceeding 0.95 of increased treatment log-odds, compared to 2 adverse events flagged by a standard analysis.

5. Extended Methods – Interim Analyses

The Berry and Berry model may be extended for use at interim analyses by dividing the trial duration into intervals and considering both the time in study of the patients and the number of adverse events that occur over each interval of the trial. If the trial is split into *H* intervals, with *B* SOCs, k_b adverse events in SOC *b*, and *C* different covariate patterns among the data, then the data model is:

$$X^{(c)}_{bj,h} \sim \text{Poisson}(\lambda^{(c)}_{bj,h} T^{(c)}_{bj,h}) \qquad c = 1, \ldots, C$$

$$T^{(c)}_{bj,h} = \sum_{i \in \mathcal{R}^{(c)}_{bj,h}} t_{ih} \qquad h = 1, \ldots, H$$

$$b = 1, \ldots, B$$

$$\log \lambda^{(c)}_{bj,h} = \gamma_{bj,h} + x_{(c)} \theta_{bj,h} \qquad j = 1, \ldots, k_b$$

where $\mathcal{R}^{(c)}_{bj,h}$ is the set of patients with covariate pattern *c* at risk of the j^{th} adverse event in SOC *b* at the start of interval *h,* and t_{ih} is the time patient *i* spends in interval *h*.

6. Software Implementation

All of the methods are implemented in the R package c212 (https://CRAN.R-project.org/package=c212).

References

• Scott M. Berry and Donald A. Berry. Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model. Biometrics, 60(2):418–426, 2004.
• D. V. Mehrotra and A. J. Adewale. Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals. Stat Med, 31(18):1918–30, 2012.
• J. X. Hu, H. Zhao, and H. H. Zhou. False discovery rate control with groups. J Am Stat Assoc, 105(491):1215–1227, 2010.