Type-1 Error Control in Safety Signal Detection in Clinical Trial Data

A Talk from the ASA Biopharm Safety Monitoring Working Group

August 2, 2017 Baltimore MD

Authors: Michael Fries, Judy Li and Matilda Sanchez-Kam

Safety Monitoring Working Group is sponsored by Biopharmaceutical Section of the American Statistical Association

Is inference and type 1 error control as important for safety monitoring?

Choices of action

- **Increase target monitoring**
- **Dose modification**
- **Patient selection**
- **Halt enrollment**
- **Halt trial**
- **SUSAR, other reporting**
- **Identify hypothesis**

- What level of evidence is required to make a choice?
- Is the same level of evidence required for all adverse events?
- Type-1 error has different implications

Type II error may be considered more important in safety setting (Xia 2011).

ICH-E9: "Statistical adjustments for multiplicity to quantify the Type I error [false positive] are appropriate, but the Type II error [false negative] is usually of more concern."

BIOPHARMACEUTICAL SECTION

SPERT Recommendations (Crowe, 2009)

Control type-1 error and understand level of evidence through good statistical practice

Tier I AE: pre-specified detailed analysis and hypothesis testing

Tier 2 AE: signal detection among common events

Tier 3 AE: Descriptive analysis of infrequent AEs

BIOPHARMACEUTICAL SECTION

Detecting signals in adverse events

Statistical Challenge: Multiplicity

1. Many adverse events, with different implications
2. Monitoring over time
3. Possibly, several arms

How can we draw inferences in the face of controlling, or at least understanding, type 1 error?

False Discovery Rate (FDR) as a useful measure of type-1 error in this setting

BIOPHARMACEUTICAL SECTION

Approaches – Non-inferential

Don’t apply statistics. Just summarize the incidence and use clinical judgment.

Provide aids to clinical judgment:

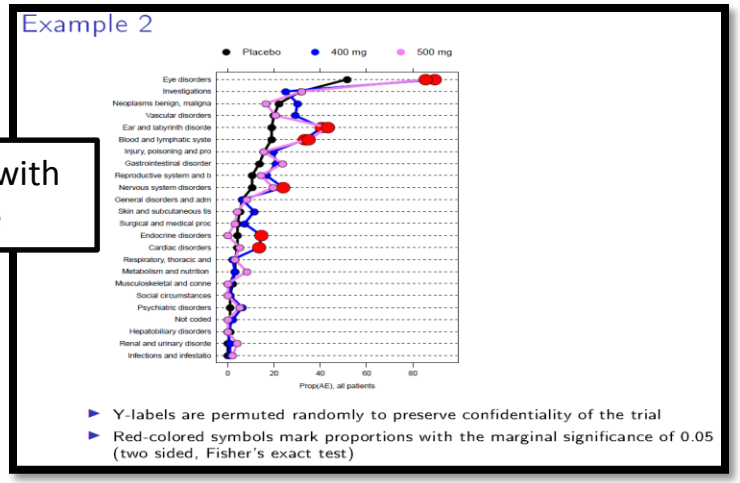
- Apply statistical tests, but consider them flagging exercises as “possible” signals
- Estimating risk difference, risk ratios, or odds ratios
- Graphical Tools

BIOPHARMACEUTICAL SECTION

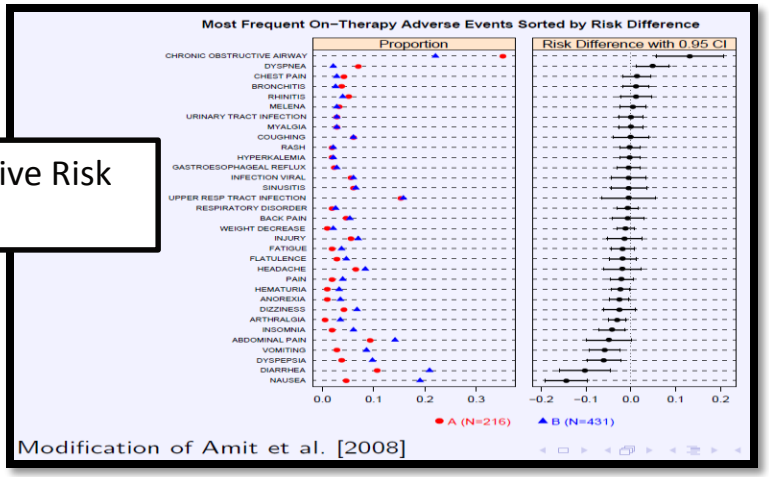
Examples of Graphs for Safety Monitoring

TIER 1 Event Visuals

Dot plot with inference

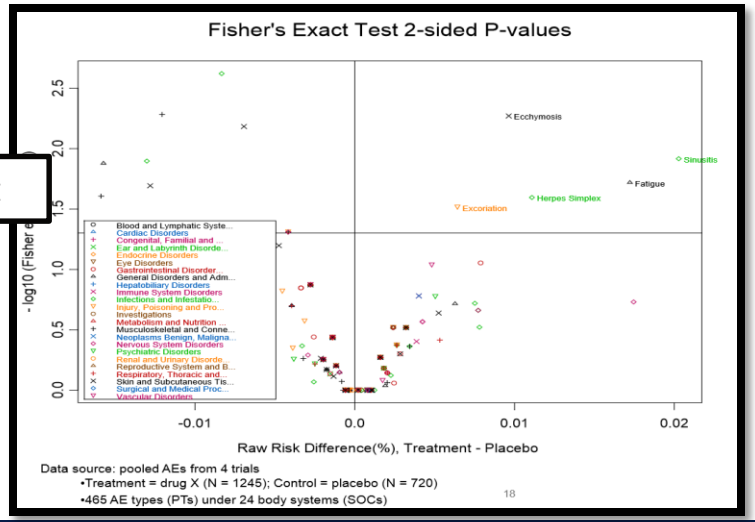


Relative Risk Plot

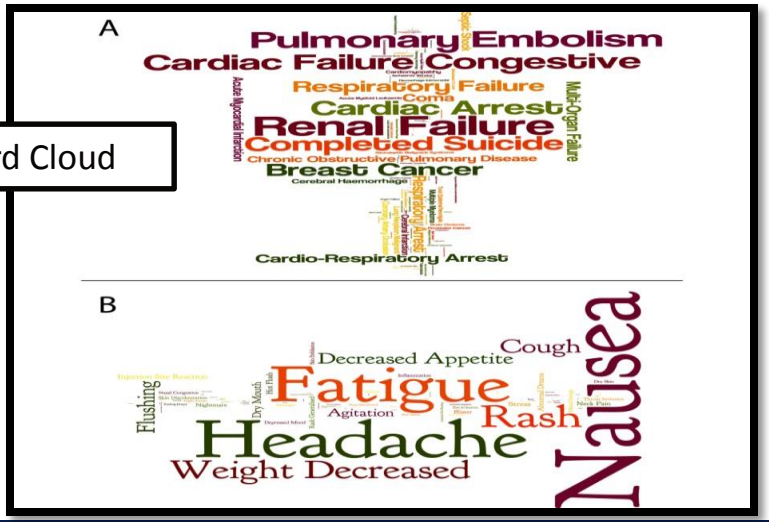


TIER 2 Event Visuals

Volcano Plot



Word Cloud



These techniques may be useful.

They do not address multiplicity and/or they do not allow inference

What are more considered approaches to Signal Detection?

BIOPHARMACEUTICAL SECTION

Controlling multiplicity from AEs

New Double False Discovery Rate (Mehrotra and Adewale 2012)

Bayesian Hierarchal model (with extensions) (Berry and Berry 2004, Xia 2011)

BIOPHARMACEUTICAL SECTION

New Double FDR (Mehrotra and Adewale, 2012) (1 of 2)

Method for screening frequent (TIER 2) adverse events, controlling for multiplicity, adjustment to (original) double FDR

Uses hierarchal approach with SOC (Body system)

Flexible adjustment for any analysis of AEs that create p-values.

Relatively simple to implement

BIOPHARMACEUTICAL SECTION

Flagging Tier 2 AEs: New DFDR Method
Mehrotra & Adewale (Statistics in Medicine, 2012)

- Apply FDR adjustment to p_{ij} $(1 \le j \le m_i)$ in body system i
  Let $p_i^* =$ smallest $\tilde{p}_{ij}$ in body system i

- **Step 1: flag body system(s)**
  - ◊ Apply FDR-adjustment to p_i^* $(1 \le i \le s)$
    Let $\tilde{p}_i^* =$ FDR-adjusted p_i^*
  - ◊ Flag body system i if $\tilde{p}_i^* < \alpha$

  Step 2: flag AEs within flagged body system(s)
  - ◊ Pool the p-values from the flagged body systems into a single family $F \equiv \{p_{i,j} \mid \tilde{p}_i^* < \alpha\}$
  - ◊ Apply FDR adjustment to F
    Let $\tilde{p}_{ij}^F =$ FDR-adjusted p_{ij} in F
  - ◊ Flag AE corresponding to p_{ij} if $\tilde{p}_{ij}^F < \alpha$

Slide content provided by Mehrotra, personal communication

Bayesian Hierarchical Model (1 of 2)

Berry and Berry (2004) proposed a three-level hierarchical mixed model to account for multiplicities in AE assessment

- Basic level: individual AE (PT)
- Second level: body system (SOC) which contains a number of types of possibly related AEs
- Highest level: collection of all body systems

Assume that AEs in the same body system are exchangeable and rates of AEs are more likely to be similar within than across body systems

Decision is based on the posterior probability that the event rate on treatment is greater than on the control

BIOPHARMACEUTICAL SECTION

11

Bayesian Hierarchical Model (2 of 2)

Xia, Carlin, Ma (2011) extended BB model to the Poisson case and give guidance on decision rules

Xia, Carlin and Ma (2011) paper also provides guidance on how to choose a signal detection threshold to achieve a fair balance between false positive error rates and false negative error rates via simulation study

Bayesian methods may not require distinction between tier 2 and tier 3 events (Xia 2011)

The level of evidence required for identifying a signal can be pre-specified at the level of preferred term.

BIOPHARMACEUTICAL SECTION

Simulation Results

Xia (2011) show improvements in both power and FDR in hierarchal models (DFDR and 3-level Bayesian) over simpler models (did not include new DFDR)

Mehrotra (2012) showed new double FDR had superior type-1 error control to double FDR with only a slight loss in power (did not compare to Bayesian methods)

Chen (2015) presents extensive simulation comparison that includes additional methodologies. These suggest

- Both new double FDR and 3 level Bayesian show good balance between FDR and sensitivity
- 3-level Bayesian's FDR changes with sample size, and can be unstable at low sizes (assuming modest differences in AE rates)

BIOPHARMACEUTICAL SECTION

These methods have been around for a while – but why aren't we using them?

BIOPHARMACEUTICAL SECTION

Specific Challenges in Assessing Operating Characteristics

Prior to study start, we don't know how large the multiplicity problem is

Level of evidence required to act on specific adverse events may be different (cancer vs headache) and would need to be established

Safety signals may arise from combinations of AEs, rather than a single AE.

Simulations with Bayesian Hierarchical Model is challenging, run time can be longer, and assessing convergence within simulation setting.

Time available: already pressed to do more sophisticated work on efficacy side.

BIOPHARMACEUTICAL SECTION

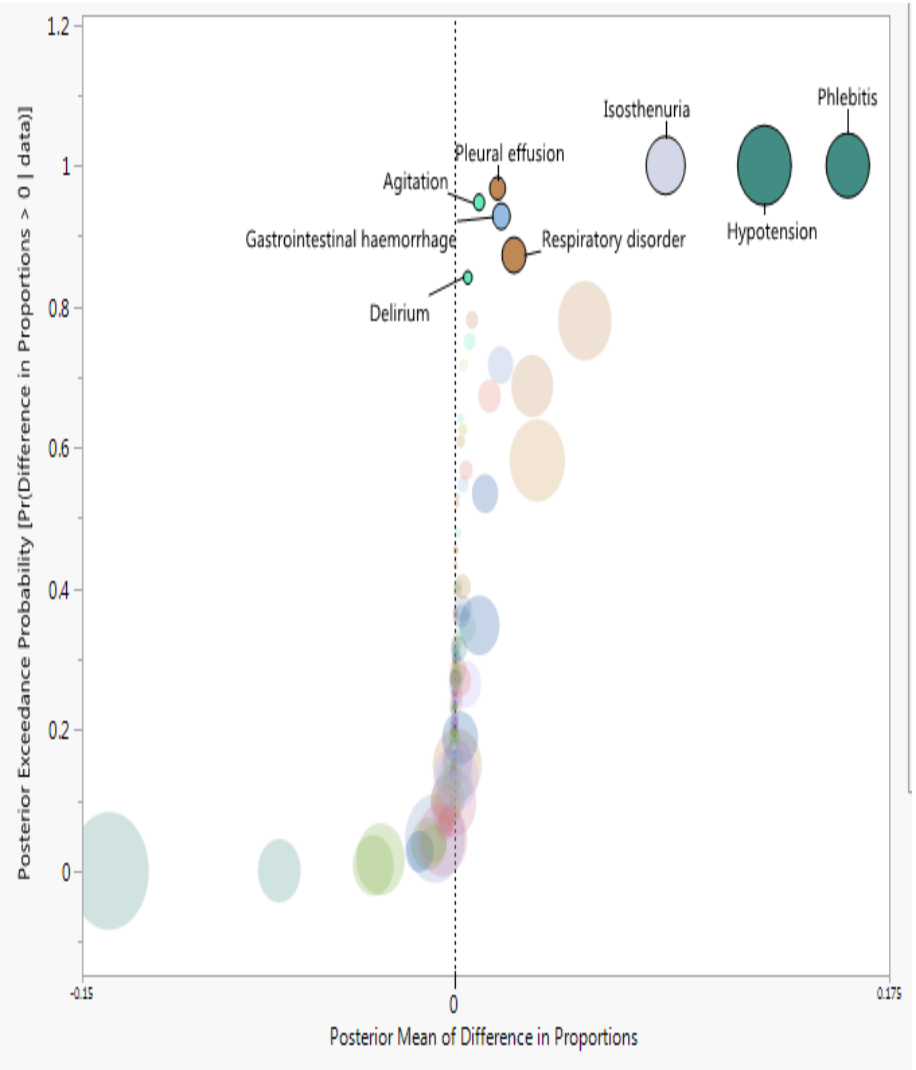
Ease of Use – Access to software

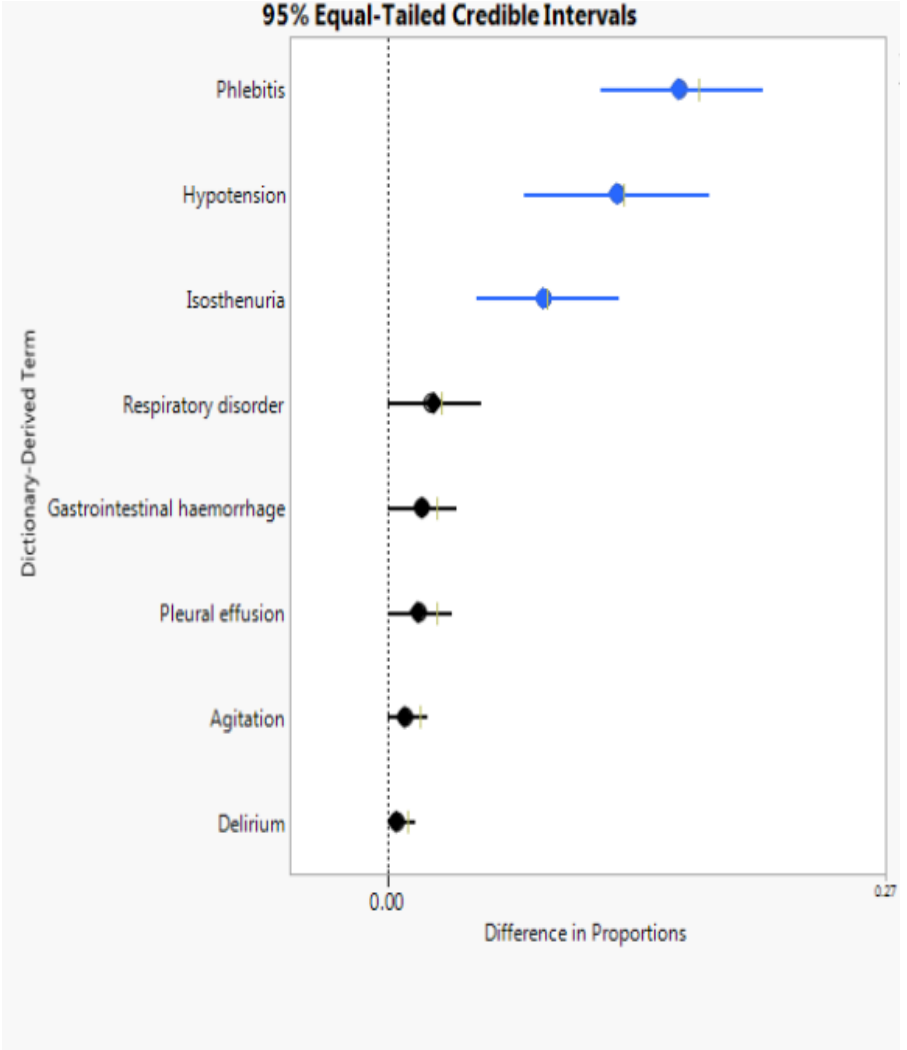
- R Package c212 for an implementation of the Bayesian Hierarchical model along with other methods for error control when testing multiple hypotheses: http://personal.strath.ac.uk/raymond.carragher/files/c212/c212-manual.pdf
- SAS
  - example of Bayesian Hierarchical model: http://blogs.sas.com/content/jmp/2013/04/26/analyzing-adverse-events-using-bayesian-hierarchical-models/
  - new DFDR easy with proc multtest.

Bayesian tools: STAN, BUGS family: Winbugs, Openbugs, JAGS (examples can be found on internet)

BIOPHARMACEUTICAL SECTION

Interpretability





Zinc, private communication

Compare and contrast 2 methods

Bayesian Hierarchal

Provides point estimates, posterior probabilities allowing credible intervals.

Flexible criteria could be implemented, for example, based on nature of AE

Harder to implement – requires level of expertise with Bayesian methods.

Different methodologies require different implementation of the model.

nDFDR

Easy to implement

Can be applied to any statistical method that provides p-values

P-values only – no confidence interval

Less clear how to adjust for nature of AE

Multiplicity Due to Repeated Monitoring

It has been argued that the “Natural” control of Bayesian methods in repeated assessments over time (likelihood principle minimizes linear combination of type 1 and type 2 error, Spiegelhalter 2004)

Chen 2013 Introduced a Bayesian method for controlling for multiple looks, implementing Bayesian decision-theoretic approach to minimize the posterior expected loss due to both type 1 and type 2 error.

We have not identified a publication of frequentist approach to multiplicity due to repeated monitoring in safety signal detection among adverse events.

BIOPHARMACEUTICAL SECTION

New DFDR example (in meta analysis)

One out of 40 AEs "flagged" prior to multiplicity adjustment based on adjusted CI for OR excluding 1 (rheumatoid arthritis). Not flagged after the pre-specified New DFDR multiplicity adjustment (details omitted).

Interpretation: no statistical evidence that vaccine increases the risk of rheumatoid arthritis.

Correct interpretation? Presumably yes, because …

Post meta-analysis note: separate trial (N ~ 4000) revealed 4 rheumatoid arthritis cases: 2 in vaccine group, 2 in placebo group.

Source: Mehrotra, personal communication

BIOPHARMACEUTICAL SECTION

Recommendations and possible next steps

Keep these methods in your toolbox!

Both of these methods offer advantages and disadvantages – choose what's best for your situation

Share experiences: we need to keep learning.

Some areas for further work:

1. Is MedDRA hierarchy the best choice? Alternatives?
2. Greater understanding of type-1 error control for repeated assessments.
3. Downgrading a signal based on accumulating data

BIOPHARMACEUTICAL SECTION

Bibliography

Amy Xia, Haijun Ma, and Bradley P. Carlin. Bayesian Hierarchical Modeling for Detecting Safety Signals in Clinical Trials. *Journal of Biopharmaceutical Statistics*, 21: 1006–1029, 2011.

International conference on harmonization of technical requirements for registration of pharmaceuticals for human use. ICH Harmonized Tripartite Guideline, Statistical Principles for Clinical Trials E9. 1998

Crowe, B., Xia, H. A., Berlin, J., Watson, D., Shi, H., Lin, S., Kuebler, J., et al. (2009). Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: A report of the Safety Planning, Evaluation and Reporting Team (SPERT). *Clinical Trials* 6:430–440.

Mehrotra, Devan V., and Adeniyi J. Adewale. "Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals." *Statistics in medicine* 31.18 (2012): 1918-1930.

Berry, S., Berry, D. (2004). Accounting for multiplicities in assessing drug safety: A threelevel hierarchical mixture model. *Biometrics* 60:418–426.

Chen, Maggie, et al. "Evaluation of statistical methods for safety signal detection: a simulation study." *Pharmaceutical statistics* 14.1 (2015): 11-19.

Chen, Wenfeng, et al. "A Bayesian Group Sequential Approach to Safety Signal Detection." *Journal of biopharmaceutical statistics* 23.1 (2013): 213-230.