

Adaptive and Repeated Cumulative Meta-Analyses for Safety Signal Detection

Hui Quan, Meehyung Cho and others

ASA Biopharmaceutical Section
Regulatory-Industry Statistics Workshop

September, 2018

Adaptive and repeated cumulative meta-analyses of safety data during a new drug development process

Hui Quan,* Yingqiu Ma, Yan Zheng, Meehyung Cho, Christelle Lorenzato, and Carole Hecquet

During a new drug development process, it is desirable to timely detect potential safety signals. For this purpose, repeated meta-analyses may be performed sequentially on accumulating safety data. Moreover, if the amount of safety data from the originally planned program is not enough to ensure adequate power to test a specific hypothesis (e.g., the noninferiority hypothesis of an event of interest), the total sample size may be increased by adding new studies to the program. Without appropriate adjustment, it is well known that the type I error rate will be inflated because of repeated analyses and sample size adjustment. In this paper, we discuss potential issues associated with adaptive and repeated cumulative meta-analyses of safety data conducted during a drug development process. We consider both frequentist and Bayesian approaches. A new drug development example is used to demonstrate the application of the methods. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: conditional power; combination test; type I error rate control; fixed and random effects models; noninferiority

Background (1/2)

- A compound was developed for treating chronic insomnia.
- The program
 - **46 studies and 5987 patients**
 - **significant effects on multiple efficacy endpoints**
 - **27 in active treatment versus 0 patients in placebo had diverticulitis.**
- After discussions/reviews with external experts, the company decided not to file the NDA.
- Several years' effort and a lot of resources were wasted.
- The question:
 - **What should be done differently?**
- Intuitive thinking
 - **Repeated meta-analyses**

Background (2/2)

- Meta analysis associates with
 - **A powerful tool for synthesizing information**
 - **The chance of detecting signal – usually impossible with data of individual studies**
 - **Much more precise estimate of effect**
 - Meta analysis is typically performed only once after the completions of all studies
 - **No issue of type I error rate inflation**
 - **No need for adjusting the inference statistic**
 - If repeated meta analyses are conducted on cumulative data and adaptations on sample sizes will be performed for adaptive objectives
 - **Adjustment is necessary**
 - **We want to conduct a systematic review on related issues and methodologies**
-

Data Source

- The Clinical Development Plan (CDP) outlines the number of studies, the populations, the objectives, the sample sizes and timelines of study completions
 - Data are accumulated over time as study is completed one after another.
 - The team can plan repeated meta-analyses whenever there are substantial additional safety data – from a group of small studies or a major phase III study.
 - Stratified analysis is usually used particularly if some studies used imbalanced designs.
 - The number ***K*** of analyses should be pre-specified
 - Data of ongoing studies should be monitored by DMC and not be included in repeated meta-analyses
 - Updated safety results should be shared with DMCs
-

Hypotheses

- Inferiority assessment (IA or signal detection) hypothesis

$$H_0 : \delta = 0 \text{ versus } H_a : \delta > 0$$

- Exploratory in nature
- Lack of power for rare events
- Multiplicity – when there is no focus

- Non-inferiority hypothesis – more formal

$$H_0' : \delta \geq \Delta \text{ versus } H_a' : \delta < \Delta$$

- Primary focus of safety analysis – e.g., FDA guideline for Type 2 diabetic drugs on CV events
- Enough data for enough power – conditional power calculation during the process
- $\Delta=0$ is for superiority assessment

- Adaptation between the two hypotheses

Type I error rate control

- Repeated analyses will inflate Type I error rate for IA if no adjustment is performed.
 - **To kill a drug with a large chance.**
- If Type I error rate is controlled under a stringent rule, a toxic drug could be claimed as a safe drug.
- Strict Type I error rate control for IA may not be necessary. Nevertheless, knowing the Type I error rate of making a claim will help us to make wise decision
- Adjustment for Type I error rate control will be considered.

Methods (1/5)

- P-value combination (Fisher): $k=1, \dots, K$

$$T_k = -2 \log(P_1 P_2 \dots P_k) \sim \chi_{2k}^2$$

- Equal weight for each p-value
- For given alpha spending function $\alpha(k)$, significant effect could be declared early with critical value c_k

$$\Pr(T_1 < c_1, \dots, T_{k-1} < c_{k-1}, T_k \geq c_k \mid H_0) = \alpha(k) - \alpha(k-1)$$

- The derivation of critical values based on multivariate chi-square distribution may not be easy
- Studies could be added or cancelled during the development process --- sample size adaptation

Methods (2/5)

- Take the case of $K=2$ for example,
 - If $P_1 > 0.2$, only the originally planned second study is conducted with p-value= P_2
 - If $0 < P_1 \leq 0.2$, one additional study will be conducted with p-value= Q_2
 - Under H_0 , the naïve test statistic
$$-2\log[P_1(P_2I[P_1 > 0.2])(P_2Q_2I[P_1 \leq 0.2])]$$
will not follow χ_4^2 or χ_6^2 and is not a valid test statistic.
 - A right statistic is
$$-2\log(P_1P_2^*) \sim \chi_4^2$$
 - where $P_2^* = P_2$ if $P_1 > 0.2$ and otherwise P_2^* is the p-value from
$$\Pr(\chi_4^2 \geq -2\log(P_2Q_2))$$

Methods (3/5)

- Weighted combination test:
 - For pre-specified weights, under H_0

$$V_k = \sum_{i=1}^k w_i Z_i / \sqrt{\sum_{i=1}^k w_i^2} \sim N(0,1)$$

- Flexible for NI assessment, estimation, conditional power calculation and Type I error rate control

Methods (4/5)

- Estimation of treatment effect

$$\hat{\delta}_i \sim N(\delta, \sigma_i^2) \text{ then } Z_i' = (\hat{\delta}_i - \delta) / \hat{\sigma}_i \sim N(0,1)$$

$$T_k' = \left(\sum_{i=1}^k w_i Z_i' \right) / \sqrt{\sum_{i=1}^k w_i^2} \sim N(0,1)$$

The estimate of treatment effect at the k th analysis

$$\hat{\delta}^{(k)} = \left(\sum_{i=1}^k w_i \hat{\delta}_i / \hat{\sigma}_i \right) / \left(\sum_{i=1}^k w_i / \hat{\sigma}_i \right)$$

The upper bound

$$UB^{(k)} = \left(\sum_{i=1}^k w_i \hat{\delta}_i / \hat{\sigma}_i + c_k' \sqrt{\sum_{i=1}^k w_i^2} \right) / \left(\sum_{i=1}^k w_i / \hat{\sigma}_i \right)$$

Methods (5/5)

- Reject H_0' and claim the non-inferiority if

$$UB^{(k)} \leq \Delta$$

- Note that

- **Even the NI can be claimed early, the sponsor will not stop the ongoing trials and will initiate other trials according to the original CDP.**
- **All alpha should be reserved for the final analysis to increase power.**
- **If potentially there will not be enough power for NI, studies may be added to increase the total sample size based on conditional power.**

Power comparison (1/2)

- Lan et al. (2003) propose to apply the law of iterated logarithm to repeated meta-analyses
 - **No need to pre-specified the number of analyses.**
 - **To ‘penalize’ the Z value to account for multiple tests by using a γ that is obtained through simulation**
 - **The test statistics are compared to the nominal value**
- Power comparison between the iterated logarithm and the weighted approach could be performed.

Power comparison (2/2)

Alpha spending $\alpha(k) = 1 - \Phi(z_{1-\alpha} / \sqrt{k/K})$ for weighted method

| δ | 10 studies | | Adding 1 study | | Adding 2 studies | |
|----------|------------|-------|----------------|-------|------------------|-------|
| | Iterated | Wted | Iterated | Wted | Iterated | Wted |
| 0.00 | 0.024 | 0.025 | 0.025 | 0.025 | 0.026 | 0.025 |
| 0.02 | 0.054 | 0.071 | 0.059 | 0.074 | 0.063 | 0.079 |
| 0.04 | 0.117 | 0.168 | 0.133 | 0.181 | 0.148 | 0.200 |
| 0.06 | 0.232 | 0.328 | 0.268 | 0.355 | 0.305 | 0.397 |
| 0.08 | 0.398 | 0.526 | 0.459 | 0.571 | 0.518 | 0.629 |
| 0.10 | 0.591 | 0.724 | 0.668 | 0.771 | 0.735 | 0.824 |
| 0.12 | 0.770 | 0.869 | 0.839 | 0.904 | 0.889 | 0.937 |
| 0.14 | 0.895 | 0.950 | 0.939 | 0.969 | 0.965 | 0.983 |

Bayesian approach

- It is a very natural and useful tool for synthesizing cumulative information in a sequential way.
 - **It is the same no matter whether the analysis is based on the updated prior distribution and the newly available data or the original prior distribution and all available data.**

$$f(\theta | x_1, x_2, \dots, x_k) \propto f(x_k | \theta) f(\theta | x_1, \dots, x_{k-1}) \propto \prod_{i=1}^k f(x_i | \theta) f(\theta)$$

- **Just perform the regular analysis with the cumulative data, particularly helpful with MCMC approach**
- Hierarchical model for binary endpoint

Diverticulitis events in the program

- The original program did not have a plan for repeated meta-analyses. ($p=1.37 \times 10^{-5}$ for pooled analysis)
- Data were used for illustration purpose.

| | Control | | Treatment | |
|---|---------|--------|-----------|--------|
| | N | Events | N | Events |
| 4 Phase I-II studies | 164 | 0 | 321 | 1 |
| 4 Phase II studies | 313 | 0 | 707 | 3 |
| Phase III study 1 | 311 | 0 | 297 | 1 |
| Phase III study 2 | 295 | 0 | 850 | 14 |
| Phase III study 3 | 345 | 0 | 617 | 8 |
| Total | 1428 | 0 | 2792 | 27 |
| Two more Phase III studies on low risk patients | | | | |

Results

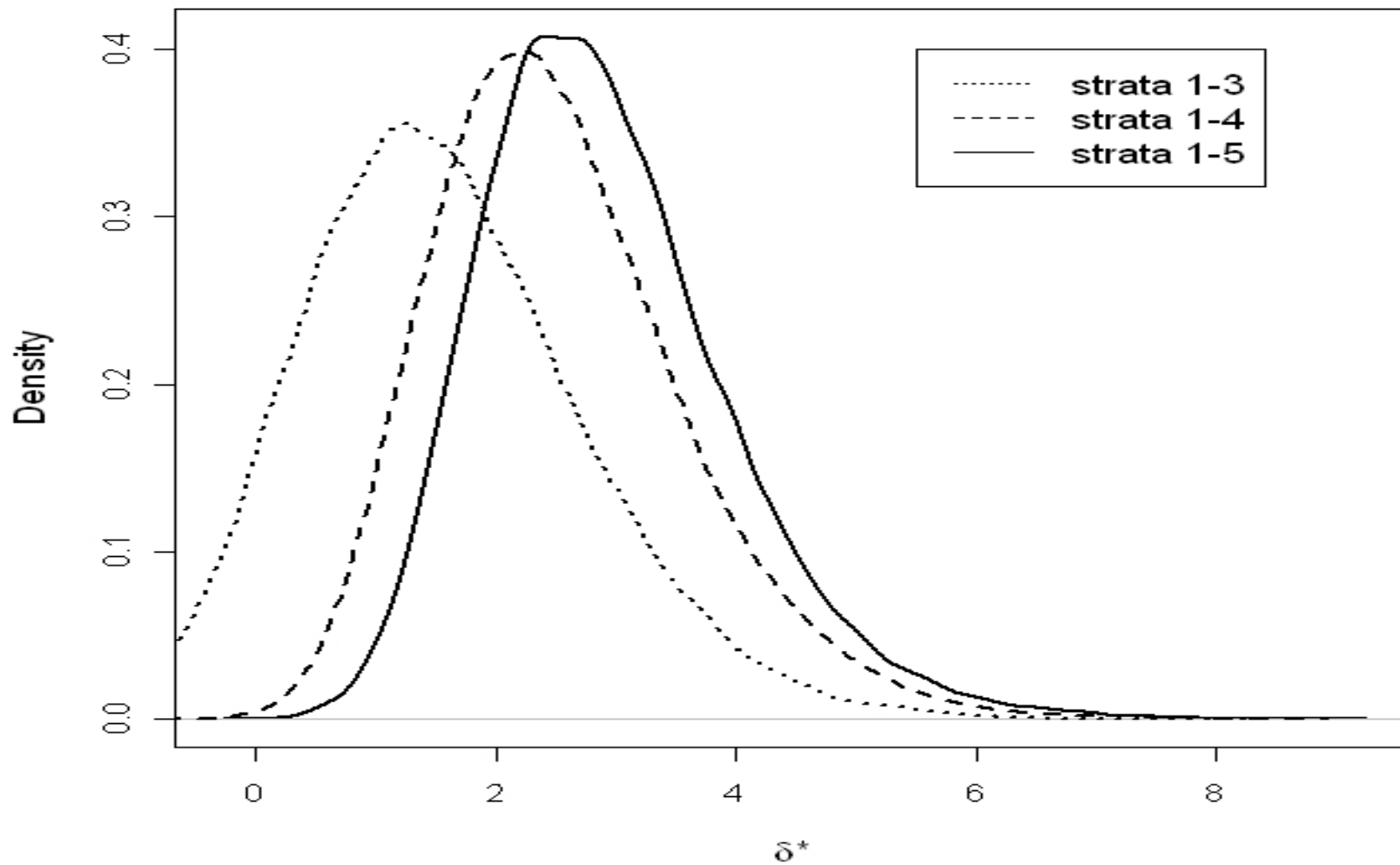
Analysis results of diverticulitis events

| Method | Outcome |
|---|--|
| P-value combination: exact test for individual strata | p-value=0.029: significant at 5 th analysis when $\alpha = 0.05$ but not 0.025. |
| Weighted combination test for difference of rates | significant at 3 rd analysis when $\alpha = 0.05$ and at 4 th analysis when $\alpha = 0.025$. |
| Weighted combination test for log risk ratio or odds ratio (0.2 continuity) | significant at 5 th analysis when $\alpha = 0.025$ and 0.05. |
| Weighted combination test for log hazard ratio (0.2 continuity) | significant at the 5 th analysis when $\alpha = 0.05$. |

Hierarchical Bayesian Model for log odds ratio

| Strata | mean | 2.5% | median | 97.5% |
|--------|-------|---------|--------|-------|
| 1-3 | 1.563 | -0.4688 | 1.468 | 4.145 |
| 1-4 | 2.522 | 0.7727 | 2.404 | 4.947 |
| 1-5 | 2.903 | 1.219 | 2.784 | 5.282 |

Posterior densities for log odds ratio on diverticulitis events



Discussion

- It is unethical for a sponsor not to closely monitor cumulative safety data during the drug development process.
 - Sponsor surely wants to timely terminate the development of an unsafe drug -- a way of efficient new drug development and saving resources!
 - Repeated meta-analyses should be planned proactively and early.
 - Type I error rate control is more important for testing the non-inferiority hypothesis than the inferiority hypothesis.
 - Showing nominal significant effect → starting point for more data → more formal analysis: NI assessment → thorough benefit/risk assessment.
 - Documentation is important in case of future legal issue.
-

References

- Der Simonian R and Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; 7: 177-188.
 - Lan KKG, Hu M and Cappelleri JC. Applying the law of iterated logarithm to cumulative meta-analysis of a continuous endpoint. *Statistica Sinica* 2003; 13: 1135-1145.
 - Hu M, Cappelleri JC and Lan KKG. Applying the law of iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes. *Clinical Trials* 2007; 4: 329-340.
 - Jennison C and Turnbull BW. Meta-analyses and adaptive group sequential designs in the clinical development process. *Journal of Biopharmaceutical Statistics* 2005; 15: 537-558.
 - FDA, Guidance for Industry: diabetes mellitus-evaluating cardiovascular risk in new antidiabetic therapies to treat Type 2 diabetes, December 2008.
 - Fisher RA. *Statistical Methods for Research Workers* 4th e. London: Oliver and Boyd 1932.
 - Berry SM and Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics* 2004; 60: 418-426.
 - Quan H, Ma YQ, Zheng Y, Cho M, Lorenzato C, Hecquet C. Adaptive and Repeated Cumulative Meta-Analyses for Safety Signal Detection during a New Drug Development Process. *Pharmaceutical Statistics* 2015.
-

Thank you