Meta-analysis of rare events in drug safety studies: A unifying framework for exact inferences

Dungang Liu

University of Cincinnati Lindner College of Business
dungang.liu@uc.edu

The 3rd Annual ASA New Jersey Chapter / Bayer Statistics Workshop
Whippany, NJ
November 13, 2016
Outline

For meta-analysis of rare events, I will present

Part I  An exact approach by combining $p$-value functions

Part II  A unifying framework for exact inferences
Part I  An exact approach by combining $p$-value functions
A Motivating Example


- Performed a meta-analysis of 48 independent clinical trials.
- Examine if the diabetes drug *Avandia* is associated with some *adverse events* (e.g. myocardial infarction).

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x_i$</td>
<td>$n_i$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>357</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>391</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>774</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>47</td>
<td>0</td>
<td>196</td>
</tr>
<tr>
<td>48</td>
<td>0</td>
<td>676</td>
</tr>
</tbody>
</table>
Result and impact of Nissen and Wolski’s meta-analysis

Result: common odds ratio (OR) (Peto’s method)

- 95% confidence interval = (1.03 1.98);
- \( P \)-value = 0.03 for testing \( H_0 : \text{OR} = 1 \ vs \ H_1 : \text{OR} \neq 1 \).

Conclusion: Avandia is significantly associated with myocardial infarction.

“Significant” impacts:

- The stock price of GlaxoSmithKline (GSK) dropped 7.8% on a single day.
- An alert issued by FDA immediately.
- Over 1000 lawsuits against GSK.
- US sales in 2Q 2007 dropped 22% compared to 2006.
Controversy 1: zero total event studies

A zero total event study refers to a study that does not observe any event in both treatment and control arms.

Avandia data: 10 zero total event studies out of 48 studies.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x_i$</td>
<td>$n_i$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>357</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>391</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>774</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>47</td>
<td>0</td>
<td>196</td>
</tr>
<tr>
<td>48</td>
<td>0</td>
<td>676</td>
</tr>
</tbody>
</table>
Controversy 1: zero total event studies

Popular meta-analysis methods:
- Standard inverse-variance method
- Mantel-Haenszel (MH) method
- Peto’s method (used by Nissen and Wolski, 2007)

How do these methods handle zero total event studies?
- Exclude from analysis (Nissen and Wolski, 2007);
- Add 0.5 to zero cells.

Both have undesirable impact on inference.

Question 1: Can we use all available data without artificially assigning numbers to zero events?
Controversy 2: asymptotics for rare events?

Empirical coverage probability of 95% confidence interval
– based on simulated data similar to Avandia data

Question 2: Can we develop a general exact meta-analysis framework for discrete data?
Nissen and Wolski’s study raised

**Question 1:** Can we use all available data without artificially assigning numbers to zero events?

**Question 2:** Can we develop a general exact meta-analysis framework for discrete data?

**Our answer:** Yes!
Problem setting

Consider $K$ independent trials with treatment and control.

- $X_i \sim \text{Binom}(n_i, \pi_{1i})$ and $Y_i \sim \text{Binom}(m_i, \pi_{0i})$.
- The odds ratio

$$\psi = \frac{\pi_{1i}/(1 - \pi_{1i})}{\pi_{0i}/(1 - \pi_{0i})} \quad (i = 1, \ldots, K)$$

is often assumed being constant across the studies.

- Question: how to infer $\psi$ when $\pi_{1i}$ and $\pi_{0i}$ are extremely low.

$$\hat{\psi}_i = \frac{x_i/(n_i - x_i)}{y_i/(m_i - y_i)} \quad (i = 1, \ldots, K)$$

Note: our following discussion applies to inference for a general parameter $\psi$ in discrete data.
Meta-analysis principles

- Conventional principle – combining point estimates

\[ \hat{\psi}_c = \frac{1}{\sum_{i=1}^{K} w_i} (w_1 \hat{\psi}_1 + \cdots + w_K \hat{\psi}_K) \]

- Our new approach – combining p-value functions

\[ p_c(\psi) = \Phi \left( \frac{1}{\sqrt{\sum_{i=1}^{K} w_i^2}} \left( w_1 \Phi^{-1}(p_1(\psi)) + \cdots + w_K \Phi^{-1}(p_K(\psi)) \right) \right) \]

- What is a p-value function?
- Why do we use such a combining formula?
What is a $p$-value function?

To make inference for $\psi$, we consider testing the hypothesis

$$H_0 : \psi = \psi^* \text{ versus } H_1 : \psi > \psi^*,$$

where $\psi^*$ is an arbitrary but fixed value on the parameter space.

We suppose that

• a $p$-value $p_i(\psi^*; X_i, Y_i)$ can be obtained based on an exact test from the $i$-th study.

Example: Using the mid-$p$ adaption of Fisher’s exact test, we obtain a $p$-value for the odds ratio:

$$p_i(\psi^*; x_i, y_i) = \text{pr}_{\psi^*}(X_i > x_i \mid T_i = t_i) + \frac{1}{2}\text{pr}_{\psi^*}(X_i = x_i \mid T_i = t_i).$$

where $X_i$ follows the noncentral hypergeometric distribution conditional on $T_i = X_i + Y_i = t_i$. 
Remarks on the \( p \)-value function

- The \( p \)-value \( p_i(\psi^*; X_i, Y_i) \) is a function defined on both the parameter space and sample space.

- Given the sample \((x_i, y_i)\), the function \( p_i(\cdot; x_i, y_i) \) is typically a distribution function on the parameter space.

- This distribution function \( p_i(\cdot) \) is called a \textit{p-value function} or a \textit{significance function} (Fraser, 1991, \textit{JASA}).

- The \( p \)-value function \( p_i(\cdot) \) can be viewed as a “distribution estimate” of the unknown parameter.

  - Singh et al., 2005, \textit{Ann. Statist.}
  - Xie et al., 2011, \textit{JASA}
  - Fraser, 2011, \textit{Statist. Sci.}
**$P$-value function curves**

**Black** solid curve – an individual $p$-value function
- Obtained by using the mid-$p$ adaptation of Fisher exact test on odds ratio in a study that observes $x_i = 1$ and $y_i = 3$ with sample sizes $(n_i, m_i) = (15, 60)$.

**Red** dashed curve – the combined $p$-value function
- Obtained by combining two independent copies of the above individual $p$-value function.
How to combine $p$-value functions?

$$p_c(\psi) = \Phi \left( \frac{1}{\sqrt{\sum_{i=1}^{K} w_i^2}} \left( w_1 \Phi^{-1}(p_1(\psi)) + \cdots + w_K \Phi^{-1}(p_K(\psi)) \right) \right)$$

- $p_1(\psi), \ldots, p_K(\psi)$ are $p$-value functions.
- $\Phi(\cdot)$ is the CDF of the standard normal distribution.
- $w_i$ is the weight assigned to the $i$-th study.

The simple combining formula yields statements that explicitly account for the impact of individual studies on the overall inference (e.g., efficiency/power, type I error rate).
Inference from a \( p \)-value function

The combined \( p \)-value function \( p_c(\psi) \) can be used for making inference for the parameter \( \psi \).

- **Point estimate.** The median of the distribution \( p_c(\psi) \), namely \( \hat{\psi}_c = p_c^{-1}(1/2) \), can be used as a point estimator.

- **Interval estimate.** The interval \((p_c^{-1}(\alpha/2), p_c^{-1}(1 - \alpha/2))\) can be used as a 100\((1 - \alpha)\)% confidence interval.

- **\( P \)-value.** The value of \( p_c(\psi^*) \) can be readily used as the overall \( p \)-value for testing the hypothesis \( H_0 : \psi = \psi^* \) vs \( H_1 : \psi > \psi^* \).
Figure 1. The plot is a graphical illustration on making inference using a confidence distribution, including examples of point estimators (mode $\hat{\theta}$, median $M_n$ and mean $\bar{\theta}$), a level 95% confidence interval and a one-sided $p$-value.
Preview of desirable properties

- **Practical usefulness**
  - Our approach includes in the analysis all available data, including zero total event studies, without using any artificial correction for zero event.

- **Methodological broadness**
  - Our framework encompasses a broad class of exact meta-analysis methods, as it permits broad choices for the combining elements, such as tests used in individual studies, and any parameter of interest.

- **Theoretical soundness**
  - Our approach yields statements that explicitly account for the impact of individual studies on the overall inference (efficiency/power and type I error rate).

- **Numerical superiority**
  - Our approach outperforms existing commonly used meta-analysis methods in the setting of rare events.
Data sets used in numerical studies

1. Avandia data (Nissen and Wolski, 2007, Table 3).
   - \( K = 48 \);
   - \( \text{median}(n_i) = 222, \text{median}(m_i) = 142 \).
   - 10 zero total event studies.

2. Promotion data (Gastwirth, 1984, Table 8).
   - \( K = 10 \);
   - \( \text{median}(n_i) = 25, \text{median}(m_i) = 9 \).
   - Zero events in one arm across all studies.
## Promotion data

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Whites</th>
<th></th>
<th>Blacks</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x_i$</td>
<td>$n_i$</td>
<td>$y_i$</td>
<td>$m_i$</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>20</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>17</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>15</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>30</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>31</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>31</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>30</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>34</td>
<td>0</td>
<td>13</td>
</tr>
</tbody>
</table>
Simulation setting

We mimic the data structure of Avandia data or promotion data.

- $K = 48 \times 2 \times 2$ tables with the row margins being the same as the tables in Avandia data (or promotion data, $K = 10$).
- A fixed odds ratio (OR) ranging from 1 to 10.
- $\{\pi_{0i}, i = 1, 2, \ldots, 48\}$ are generated from $U(0, \xi)$, where $\xi = 0.01, 0.05, 0.1$.
- $\{\pi_{1i}, i = 1, 2, \ldots, 48\}$ are determined by

\[
\text{logit}(\pi_{1i}) = \log(\text{OR}) + \text{logit}(\pi_{0i})
\]
Methods for numerical comparison

- Combining the $p$-value functions;
- Combining the beta-adjusted $p$-value functions;
- MH method without CCs (excluding zero total event studies);
- MH method with 0.5 CCs for zero events;
- Peto’s method without CCs (excluding zero total event studies);
- Peto’s method with 0.5 CCs for zero events.
Part I: LLX’s exact approach

Part II: Tian et al.’s exact approach

References
# Real data analysis

<table>
<thead>
<tr>
<th>Avandia data</th>
<th>Promotion data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>95% CI</strong></td>
<td><strong>95% CI</strong></td>
</tr>
<tr>
<td><strong>Proposed exact</strong></td>
<td><strong>Proposed exact</strong></td>
</tr>
<tr>
<td>(0.972, 2.001)</td>
<td>(0.842, $\infty$)</td>
</tr>
<tr>
<td><strong>Proposed exact (adj)</strong></td>
<td>(1.037, 2.004)</td>
</tr>
<tr>
<td>(1.037, 2.004)</td>
<td>(1.054, $\infty$)</td>
</tr>
<tr>
<td><strong>MH</strong></td>
<td><strong>MH-CC</strong></td>
</tr>
<tr>
<td>(1.029, 1.978)</td>
<td>(0.919, 1.647)</td>
</tr>
<tr>
<td><strong>MH-CC</strong></td>
<td>(0.921, 1.659)</td>
</tr>
<tr>
<td>(0.921, 1.659)</td>
<td>(0.776, 4.270)</td>
</tr>
<tr>
<td><strong>Peto</strong></td>
<td><strong>Peto-CC</strong></td>
</tr>
<tr>
<td>0.071</td>
<td>0.080</td>
</tr>
<tr>
<td><strong>Peto</strong></td>
<td>(1.031, 1.979)</td>
</tr>
<tr>
<td>0.029</td>
<td>(1.054, $\infty$)</td>
</tr>
<tr>
<td><strong>Peto-CC</strong></td>
<td>(1.522, 12.86)</td>
</tr>
<tr>
<td>0.163</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.174</td>
</tr>
<tr>
<td></td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>0.168</td>
</tr>
</tbody>
</table>

CI: confidence interval;

$P$: $p$-value for hypothesis testing $H_0 : \psi = 1$ versus $H_1 : \psi \neq 1$;

adj: apply beta adjustment to individual $p$-value functions;

CC: add 0.5 continuity corrections to zero events.
Summary

Meta-analysis for discrete data

\[ \downarrow \]

Combining point estimates

\[ \downarrow \]

Combining \textit{p}-value \textit{functions} based on exact tests
Part II  A unifying framework for exact inferences

- LLX’s method of combining $p$-value functions
- Tian et al.’s method of combining confidence intervals
  – (Tian et al., 2009, Biostatistics)