Meta-Analysis of Rare Adverse Events in Randomized Clinical Trials: Bayesian and Frequentist Methods

Gary Rosner\textsuperscript{1}  Hwanhee Hong\textsuperscript{2}  Chenguang Wang\textsuperscript{1}

\textsuperscript{1}Division of Biostatistics and Bioinformatics
Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University School of Medicine
\textsuperscript{2}Department of Biostatistics and Bioinformatics
Duke University School of Medicine

NJ Chapter of the ASA 40th Spring Symposium:
Clinical Trials for Rare Diseases and Pediatric Populations
Challenges in Design and Analysis
28 June 2019
Introduction

- Regulatory approval of a drug or device involves an assessment of benefits AND risks of adverse events associated with the therapeutic agent or device.
- For a rare adverse event, one or two RCTs may not be able to provide enough of a signal of the outcome.
- Meta-analysis pools information on events over trials (e.g., Phase 3 RCTs or surveillance studies) to increase the sample size (and statistical information).
  - Larger sample size should improve chance of seeing rare adverse effects.
- Some high profile situations have suggested increased risks of uncommon serious AEs and generated controversy.
- Many papers compare performance of many frequentist meta-analysis methods in the setting of rare events, but only a few examined Bayesian approaches.
Objective

- We summarize the current state of meta-analysis with rare events.
- We compare both frequentist and Bayesian meta-analytic approaches via extensive simulation studies.
- We will
  - Review of meta-analysis models
  - Data analysis using Nissen and Wolskis rosiglitazone data
  - Simulation study
Meta-Analysis

- Meta-analysis is a statistical technique for combining the findings from independent studies to assess the clinical effectiveness or safety of healthcare interventions.
- This approach provides a quantitative (statistical) estimate of net benefit and effect heterogeneity aggregated over all the included studies.
Vioxx (Rofecoxib) and Cardiovascular Side Effects

Risk of Cardiovascular Events Associated With Selective COX-2 Inhibitors

Debabrata Mukherjee, MD
Steven E. Nissen, MD
Eric J. Topol, MD

Atherosclerosis is a process with inflammatory features and selective cyclooxygenase 2 (COX-2) inhibitors may potentially have antiatherosclerotic effects by virtue of inhibiting inflammation. However, by decreasing vasodilatory and antiaggregatory prostacyclin production, COX-2 antagonists may lead to increased prothrombotic activity. To define the cardiovascular effects of COX-2 inhibitors when used for arthritis and musculoskeletal pain.

Merck Pulls Vioxx From Market After Link To Heart Problems

Drug’s Demise Raises Concerns About Company’s Future; Loss of $2.5 Billion in Revenue

By Barbara Martinez, Anna Wilde Mathews, Joann S. Lublin and Winslow Staff Reporters of The Wall Street Journal

Updated Oct. 1, 2004 12:01 am ET

vollently withdrew Vioxx from the market in Research published in the medical journal Lancet states that 88,000 Americans had heart attacks from Vioxx, and 38,000 of them died.

increased the risk of heart attack in some patients, and Merck was accused of hiding some of the side-effects it had seen in trials. The company maintains that it is not at fault, but its recent and unexpected announcement that it will pay almost $5 billion to settle remaining Vioxx lawsuits shows was to show that Vioxx avoided
Avandia (Rosiglitazone) and Cardiovascular Side Effects

Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes

Steven E. Nissen, M.D., and Kathy Wolski, M.P.H.

Rosiglitazone Revisited

An Updated Meta-analysis of Risk for Myocardial Infarction and Cardiovascular Mortality

Steven E. Nissen, MD; Kathy Wolski, MPH

Prescriptions
Making Sense of the Health Care Debate

F.D.A. Panel Votes to Restrict Avandia

By Gardiner Harris  July 14, 2010 8:59 am

A majority of the advisory panel of the Food and Drug Administration voted today to restrict the sales of Avandia, a controversial diabetes drug, because of its potential risk for causing heart attacks. The 33-member advisory committee was deeply divided. Twelve voted to remove Avandia from the market altogether; 10 for continued sale but with new label revisions and possible restrictions; 7 to add more warnings and 3 for no change at all. But the votes can also be viewed as a decision by a majority, 31, to continue allowing sales of Avandia, with more restrictions. A

I want the FDA to order the company to withdraw the drug. The evidence is there.

—Dr. Steven Nissen, Cleveland Clinic

Let's treat our patients instead of sensationalizing half-truths and conclusions that are not yet formed.
Another Recent Incident: Varenicline (Pfizer’s Chantix)

Varenicline: First approved nicotinic receptor partial agonist for smoking cessation

RCT in patients with CV disease (Rigotti et al. Circulation. 2010;121:221-229)

Conclusions: Varenicline effective for smoking cessation in smokers with CV disease. “It was well tolerated and did not increase cardiovascular events or mortality”
Meta-analysis around the same time
(Singh S et al. *CMAJ*. 2011;183:1359-66.)

**Risk of serious adverse cardiovascular events associated with varenicline: a systematic review and meta-analysis**

Sonal Singh MD MPH, Yoon K. Loke MBBS MD, John G. Spangler MD MPH, Curt D. Furberg MD PhD

- 14 double-blind RCTs involving 8216 participants
- Selected pubs that reported CV events as “serious adverse events associated with the use of varenicline.”
- Conclusion: “Varenicline was associated with a significantly increased risk of serious adverse cardiovascular events compared with placebo”
- 1.06% [52/4908] varenicline group v. 0.82% [27/3308] placebo group
- Peto OR 1.72, 95% CI (1.09, 2.71); I² = 0%
22 double-blind, placebo-controlled RCTs. 2 included pts w/ active CV disease, 11 enrolled participants with a history of CV disease.

Conclusion: “This meta-analysis—which included all trials published to date, focused on events occurring during drug exposure, and analysed findings using four summary estimates—found no significant increase in cardiovascular serious adverse events associated with varenicline use.”

Note: Authors warn, “For rare outcomes, summary estimates based on absolute effects are recommended and estimates based on the Peto odds ratio should be avoided.”
Network meta-analysis exploring CV AEs with pharmacotherapies for smoking cessation (Mills et al. *Circulation*. 2014;129:28-41)

10 electronic databases and accessed internal FDA reports

Included any RCT of 3 approved trts (nicotine replacement therapy, bupropion, & varenicline) that reported CV disease outcomes

63 eligible RCTs involving 21 nicotine replacement therapy RCTs, 28 bupropion RCTs, and 18 varenicline RCTs. They found

- "...no increase in the risk of all cardiovascular disease events with bupropion" (RR, 0.98; 95% CI, 0.54–1.73)
- "or varenicline (RR, 1.30; 95% CI, 0.79–2.23)

Re: major adverse CV events: "...found a protective effect with bupropion (RR, 0.45; 95% CI, 0.21–0.85) and no clear evidence of harm with varenicline (RR, 1.34; 95% CI, 0.66–2.66) or nicotine replacement therapy (RR, 1.95; 95% CI, 0.26–4.30)."

Conclusion: "Smoking cessation therapies do not appear to raise the risk of serious cardiovascular disease events."
What to Do with Meta-Analyses with Rare Events?

Issues in Meta-Analysis with Rare Events

- Sparsity of data with zero-event trials
- Insufficient statistical power in random effects models
- Potential dominance of a few large trials

Several authors argue that traditional meta-analysis methods may be ill-defined or have poor performance properties with rare events

- Shuster JJ and Walker MA. *Statistics in Medicine*, 2016;35:2467-2478
- Lane PW. *Statistical Methods in Medical Research*, 2013:22(2):117-132
- Sweeting MJ et al. *Statistics in Medicine*, 2004;23(9):1351-75
Some Alternative Meta-Analytic Methods for Rare Events

- Poisson random-effects models using a likelihood-based approach

- Combining confidence intervals

- Using an arcsine difference

A common limitation of these simulation studies

They generated simulated data sets assuming no heterogeneity in treatment effect across trials
Avandia (Rosiglitazone)

- We use the Nissen and Wolski data published in 2010.
- NW collected the data to investigate the effect of rosiglitazone on an increase in the risk of myocardial infarction (MI) or cardiovascular causes death (CV death) compared to the control group.
- The total number of trials is 56
  - 15 trials did not report any MI events.
  - 29 trials did not report any CV death events.
- Control groups in each trial vary widely.
25 of the 56 Trials

Table 1. Rosiglitazone Clinical Trials Included in the Meta-analysis

<table>
<thead>
<tr>
<th>Clinical Trial No.</th>
<th>Phase</th>
<th>No. of Weeks</th>
<th>Rosiglitazone Therapy</th>
<th>Control Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>49653/011</td>
<td>3</td>
<td>24</td>
<td>Rosiglitazone</td>
<td>357</td>
</tr>
<tr>
<td>49653/020</td>
<td>3</td>
<td>52</td>
<td>Rosiglitazone</td>
<td>391</td>
</tr>
<tr>
<td>49653/024</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone</td>
<td>774</td>
</tr>
<tr>
<td>49653/093</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone±metformin</td>
<td>213</td>
</tr>
<tr>
<td>49653/094</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone and metformin</td>
<td>232</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>1967</strong></td>
</tr>
</tbody>
</table>

**Trials Included in Original Registration Package**

**Additional Phase 2, 3, and 4 Efficacy Trials**

<table>
<thead>
<tr>
<th>Clinical Trial No.</th>
<th>Phase</th>
<th>No. of Weeks</th>
<th>Rosiglitazone Therapy</th>
<th>Control Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>100684</td>
<td>4</td>
<td>52</td>
<td>Rosiglitazone and glyburide</td>
<td>43</td>
</tr>
<tr>
<td>49653/143</td>
<td>4</td>
<td>24</td>
<td>Rosiglitazone and glyburide</td>
<td>121</td>
</tr>
<tr>
<td>49653/211</td>
<td>4</td>
<td>52</td>
<td>Rosiglitazone and usual care</td>
<td>110</td>
</tr>
<tr>
<td>49653/284</td>
<td>4</td>
<td>24</td>
<td>Rosiglitazone and metformin</td>
<td>382</td>
</tr>
<tr>
<td>712753/008</td>
<td>4</td>
<td>48</td>
<td>Rosiglitazone and metformin</td>
<td>284</td>
</tr>
<tr>
<td>AVM100264</td>
<td>4</td>
<td>52</td>
<td>Rosiglitazone and metformin</td>
<td>294</td>
</tr>
<tr>
<td>BRL 49653C/185</td>
<td>4</td>
<td>32</td>
<td>Rosiglitazone±metformin</td>
<td>563</td>
</tr>
<tr>
<td>BRL 49653/334</td>
<td>4</td>
<td>52</td>
<td>Rosiglitazone</td>
<td>278</td>
</tr>
<tr>
<td>BRL 49653/347</td>
<td>4</td>
<td>24</td>
<td>Rosiglitazone and insulin</td>
<td>418</td>
</tr>
<tr>
<td>49653/015</td>
<td>3</td>
<td>24</td>
<td>Rosiglitazone and sulfonylurea</td>
<td>395</td>
</tr>
<tr>
<td>49653/079</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone±glyburide</td>
<td>203</td>
</tr>
<tr>
<td>49653/080</td>
<td>3</td>
<td>156</td>
<td>Rosiglitazone</td>
<td>104</td>
</tr>
<tr>
<td>49653/082</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone and insulin</td>
<td>212</td>
</tr>
<tr>
<td>49653/085</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone and insulin</td>
<td>138</td>
</tr>
<tr>
<td>49653/095</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone and insulin</td>
<td>196</td>
</tr>
<tr>
<td>49653/097</td>
<td>3</td>
<td>156</td>
<td>Rosiglitazone</td>
<td>122</td>
</tr>
<tr>
<td>49653/125</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone and sulfonylurea</td>
<td>175</td>
</tr>
<tr>
<td>49653/127</td>
<td>3</td>
<td>26</td>
<td>Rosiglitazone and glyburide</td>
<td>56</td>
</tr>
<tr>
<td>49653/128</td>
<td>3</td>
<td>28</td>
<td>Rosiglitazone</td>
<td>39</td>
</tr>
<tr>
<td>49653/134</td>
<td>3</td>
<td>28</td>
<td>Rosiglitazone</td>
<td>551</td>
</tr>
<tr>
<td>49653/135</td>
<td>3</td>
<td>104</td>
<td>Rosiglitazone and glypizide</td>
<td>116</td>
</tr>
</tbody>
</table>
**FDA, EMA, WHO Classifications According to Frequency of Adverse Drug Reactions**

<table>
<thead>
<tr>
<th>Characterization</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very common</td>
<td>$\geq 10%$</td>
</tr>
<tr>
<td>Common (frequent)</td>
<td>$\geq 1%$ to $&lt; 10%$</td>
</tr>
<tr>
<td>Uncommon (infrequent)</td>
<td>$\geq 0.1%$ to $&lt; 1%$</td>
</tr>
<tr>
<td>Rare</td>
<td>$\geq 0.01%$ to $&lt; 0.1%$</td>
</tr>
<tr>
<td>Very rare</td>
<td>$&lt; 0.01%$</td>
</tr>
</tbody>
</table>
Literature Review

- We reviewed about 20 papers.
  - Most papers (Diamond et al. (2007); Friedrich et al. (2009); Bohning et al. (2015); Lane (2013)) re-analyzed the Nissen and Wolski data using various existing (mostly frequentist) meta-analysis models.
  - Sweeting et al. (2004) and Bradburn et al. (2007) show extensive simulation studies to compare many frequentist methods; datasets generated only under a fixed-effects model framework
    - Sweeting et al. (2004): MH with modification (+0.5) gives least biased results when sample sizes in treated and control groups are unbalanced.
    - Bradburn et al. (2007): Peto’s estimator provides least biased results when event rates < 0.01
- There are a few frequentist method papers
  - Cai et al. (2010) use Poisson regression
  - Tian et al. (2009) combine confidence intervals on risk difference scale
  - Shuster et al. (2007, 2016) propose a new ratio estimator approach forcing all studies to receive the same weight.
Assumptions in a Meta-Analysis

- One can make two basic model assumptions in a meta-analysis
  - **Common Treatment Effect (CTE):** treatment effects are constant across trials (a.k.a. fixed-effects model)
  - **Heterogeneous Treatment Effect (HTE):** treatment effects are heterogeneous across trials (a.k.a. random-effects model)
Methods We Compare

Frequentist

- CTE
  - Naïve
  - Peto
  - Mantel-Haenszel (MH)
  - MH-Data Modification (DM)
  - Inverse-Variance-DM
  - SGS-Unweighted
  - SGS-Weighted

- HTE
  - DerSimonian & Laird IV-DM
  - Simple Average
    - incl. between-study heterogeneity as in DerSimonian & Laird ($\tau_{DL}^2$)

Bayesian

- CTE
  - Logistic Regression
  - Beta Hyperprior

- HTE
  - Logistic Regression
  - Logistic w/ Shrinkage prior
  - Arm-based model
  - Beta Hyperprior
**Notation**

**Table:** Generic 2x2 table for study $i$

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of Events</th>
<th>No. of Nonevents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>$y_{i1}$</td>
<td>$n_{i1} - y_{i1}$</td>
<td>$n_{i1}$</td>
</tr>
<tr>
<td>Treatment</td>
<td>$y_{i2}$</td>
<td>$n_{i2} - y_{i2}$</td>
<td>$n_{i2}$</td>
</tr>
<tr>
<td>Total</td>
<td>$y_{i}$</td>
<td>$n_{i} - y_{i}$</td>
<td>$n_{i}$</td>
</tr>
</tbody>
</table>

- $i$ indexes study
- Assume outcomes follow binomial distributions

$$y_{i1} | p_{i1} \sim Bin(n_{i1}, p_{i1}) \text{ and } y_{i2} | p_{i2} \sim Bin(n_{i2}, p_{i2})$$
Data Modification (DM) for Cells with Zeros

If study has at least one cell with 0 events, add a constant (usually 0.5) to each cell in the study’s 2x2 table

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of events</th>
<th>No. of nonevents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>$y_{i1} + 0.5$</td>
<td>$n_{i1} - y_{i1} + 0.5$</td>
<td>$n_{i1} + 1$</td>
</tr>
<tr>
<td>Treated group</td>
<td>$y_{i2} + 0.5$</td>
<td>$n_{i2} - y_{i2} + 0.5$</td>
<td>$n_{i2} + 1$</td>
</tr>
<tr>
<td>Total</td>
<td>$y_{i.} + 1$</td>
<td>$n_{i.} - y_{i.} + 1$</td>
<td>$n_{i.} + 2$</td>
</tr>
</tbody>
</table>

- Adding 0.5 to each cell removes 1st-order bias
- (Cox & Snell, 1985; Bhaumik et al., 2012)

For our estimators...

- Mantel-Haenszel estimator with (MH(DM)) this modification
- Simple average (SA) adds 0.5 to each cell across all studies
- Inverse variance estimators include modification (CTE-IV(DM) and HTE-DL(DM))
## Naïve Pooling

### Table: 2x2 table after pooling across studies

<table>
<thead>
<tr>
<th></th>
<th>No. of events</th>
<th>No. of nonevents</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>$\sum_i y_i 1$</td>
<td>$\sum_i (n_i 1 - y_i 1)$</td>
<td>$\sum_i n_i 1$</td>
</tr>
<tr>
<td>Treated group</td>
<td>$\sum_i y_i 2$</td>
<td>$\sum_i (n_i 2 - y_i 2)$</td>
<td>$\sum_i n_i 2$</td>
</tr>
<tr>
<td>Total</td>
<td>$\sum_i y_i.$</td>
<td>$\sum_i (n_i. - y_i.)$</td>
<td>$\sum_i n_i.$</td>
</tr>
</tbody>
</table>

- We first estimate $\hat{p}_1 = \frac{\sum_i y_i 1}{\sum_i n_i 1}$ and $\hat{p}_2 = \frac{\sum_i y_i 2}{\sum_i n_i 2}$
- Then,

\[
\hat{LOR}_{naive} = \log \left[ \frac{\hat{p}_2 (1 - \hat{p}_1)}{\hat{p}_1 (1 - \hat{p}_2)} \right]
\]

with

\[
se(\hat{LOR}_{naive}) = \sqrt{\frac{1}{\sum_i y_i 1} + \frac{1}{\sum_i n_i 1 - y_i 1} + \frac{1}{\sum_i y_i 2} + \frac{1}{\sum_i n_i 2 - y_i 2}}
\]
Peto’s Method

The Peto estimator of LOR and its variance are

\[
\hat{LOR}_{\text{Peto}} = \frac{\sum_i O_i - \sum_i E_i}{\sum_i V_i}
\]

\[
\hat{\text{Var}}(\hat{LOR}_{\text{Peto}}) = 1/\sum_i V_i,
\]

where

- \(O_i = y_{i2}\)
- \(E_i = \frac{y_i \cdot n_{i2}}{n_i} \)
- \(V_i = \frac{y_i \cdot n_{i1} n_{i2}(n_i - y_i)}{n_i^2 (n_i - 1)}\)
- \(n_i = n_{i1} + n_{i2}\)
- \(y_i = y_{i1} + y_{i2}\)

Mantel-Haenszel Estimator

The MH method pools odds ratios across studies. The MH estimator of LOR and its approximate variance proposed by Robins et al.*:

\[
\hat{\text{LOR}}_{\text{MH}} = \log \left[ \frac{\sum_i y_{i2}(n_{i1} - y_{i1})}{\sum_i y_{i1}(n_{i2} - y_{i2})} \right]
\]

\[
\hat{\text{Var}}(\hat{\text{LOR}}_{\text{MH}}) = \left[ \frac{\sum_i P_i R_i}{2(\sum_i R_i)^2} + \frac{\sum_i (P_i S_i + Q_i R_i)}{2(\sum_i R_i)(\sum_i S_i)} + \frac{\sum_i Q_i S_i}{2(\sum_i S_i)^2} \right],
\]

where

- \( R_i = \frac{y_{i2}(n_{i1} - y_{i1})}{n_i} \),  \( S_i = \frac{y_{i1}(n_{i2} - y_{i2})}{n_i} \),
- \( P_i = \frac{y_{i2} + n_{i1} - y_{i1}}{n_i} \),  \( Q_i = \frac{y_{i1} + n_{i2} - y_{i2}}{n_i} \),

Recall: MH(DM) adds 0.5 to cells if study has a 0

DerSimonian and Laird (1986) proposed a weighted estimator.

- **Study-specific estimator**: \( \text{LOR}_i = \log \left( \frac{y_{i2}(n_{i1}-y_{i1})}{y_{i1}(n_{i2}-y_{i2})} \right) \)

- **Study-specific weights** \( w_i^* \) combine within- & between-study variances

\[
w_i^* = \frac{1}{v_i^*} = \frac{1}{w_i + \tau^2}
\]

- \( 1/w_i \) is within-study variance (i.e., variance of \( \text{LOR}_i \) for study \( i \))
- \( \tau^2 \) is between-study variance

- For **CTE**, we set \( \tau^2 = 0 \), denoted by CTE-IV(DM).
- For **HTE**, we estimate \( \tau^2 \), denoted by HTE-DL(DM).

Recall: The inverse-variance estimators add 0.5 to cells if study has a 0
Inverse-Variance Estimator (IV) cont’d

- Estimate $\tau^2$ as
  \[
  \hat{\tau}^2 = \begin{cases} 
  \frac{Q - df}{C}, & \text{if } Q > df \\
  0, & \text{if } Q \leq df
  \end{cases}
  \]
  - $df$ is the number of studies minus 1
  - $Q = \sum_i w_i (\text{LOR}_i - \hat{\text{LOR}})^2$
  - $\hat{\text{LOR}} = \frac{\sum_i w_i \text{LOR}_i}{\sum_i w_i}$
  - $C = \sum_i w_i - \frac{\sum_i w_i^2}{\sum_i w_i}$

The IV estimator and its variance:

\[
\hat{\text{LOR}}_{IV} = \frac{\sum_i w_i^* \text{LOR}_i}{\sum_i w_i^*}
\]

\[
\text{Var} (\hat{\text{LOR}}_{IV}) = \frac{1}{\sum_i w_i^*}
\]
In a series of papers, Shuster and colleagues argue that most meta-analyses use incorrect estimators.

Most MA estimation assumes *effects at random*
- Treatment effect and study design are independent
- Consequence: Assumes weights (e.g., sample size, variance, etc.) not correlated with trt effect

Alternative framework assumes *studies at random*
- Treatment effect and study design are not independent
- Consequence: Weights are correlated with treatment effect

Let \( \hat{p}_{jk} = \sum_i Y_{ijk} / n_{jk}, \ k = 1, 2; \ j = 1, \ldots, M \)

Unweighted estimator of relative risk: \( \text{RR} = \frac{\sum_j \hat{p}_{j2}}{\sum_j \hat{p}_{j1}} \)

Weighted estimator of relative risk: \( \text{RR} = \frac{\sum_j n_j \hat{p}_{j2}}{\sum_j n_j \hat{p}_{j1}} \)

Shuster et al. (cont’d): Estimators of Odds Ratio (OR)

- **Unweighted estimator of OR (SGS-Unwgt):**
  - Let $\hat{\pi}_k = \sum_j \hat{p}_{jk} / M$
    - Trt-specific estimate averaged across studies
  - $\text{SGS-Unwgt} = \frac{\hat{\pi}_2(1 - \hat{\pi}_1)}{\hat{\pi}_1(1 - \hat{\pi}_2)}$
  - SGS* give formula for standard error of $\log(\text{SGS-Unwgt})$
  - For confidence intervals: non-central $t_{M-1}(\cdot; \log(\text{SGS-Unwgt}), \text{StErr-Unwgt})$

- **Weighted estimator of OR (SGS-Wgt):**
  - Study-specific weight $= U_j = (n_{j1} + n_{j2})/2$
  - Let $A_{jk} = U_j\hat{p}_{jk}$ and $B_{jk} = U_j(1 - \hat{p}_{jk})$.
  - Let $\overline{A}_k = \sum_j A_{jk} / M$ and $\overline{B}_k = \sum_j B_{jk} / M$
  - $\text{SGS-Wgt} = \frac{\overline{A}_2\overline{B}_1}{\overline{A}_1\overline{B}_2}$
  - SGS* give formula for standard error of $\log(\text{SGS-Wgt})$
  - For confidence intervals: non-central $t_{M-2}(\cdot; \log(\text{SGS-Unwgt}), \text{StErr-Wgt})$

Simple Average (SA) Estimator

- Bhaumik et al. (2012): Compute average of study-specific LORs after adding 0.5 to all cells in each study’s 2x2 table
  - The study-specific $LOR_i$ (trt 2 vs trt 1):
    \[
    \widehat{LOR}_i = \log\left(\frac{y_{i2} + 0.5}{n_{i2} - y_{i2} + 0.5}\right) - \log\left(\frac{y_{i1} + 0.5}{n_{i1} - y_{i1} + 0.5}\right)
    \]
  - The SA estimator and its variance are defined as
    \[
    \widehat{LOR}_{SA} = \frac{\sum_i \widehat{LOR}_{i,1/2}}{M}
    \]
    \[
    \text{Var}(\widehat{LOR}_{SA}) = \frac{\sum_i \hat{\sigma}_i^2(\hat{\tau}^2)}{M^2}
    \]
    - $M$ is the number of studies
    - $\hat{\sigma}_i^2(\hat{\tau}^2) = [n_{i1}\hat{p}_{i1}(1 - \hat{p}_{i1})]^{-1} + [n_{i2}\hat{p}_{i2}(1 - \hat{p}_{i2})]^{-1} + \hat{\tau}^2$
    - $\hat{p}_{ik} = \frac{y_{ik} + 0.5}{n_{ik} + 1}$ for $k = 1, 2$
    - $\hat{\tau}^2$, inter-study var, as for IV estimator (DerSimonian & Laird)
Note: Compare Shuster’s Estimator to Simple Average

- Shuster’s estimators:
  1. Calculate study-specific risk, $\hat{p}_{ki}$, $k = 1, 2$
  2. Average them, $E[p_{ki}]$
  3. Calculate LOR

- SA
  1. Add 0.5 to all cells (treatment groups and studies)
  2. Calculate study-specific $LOR_i$ and then
  3. Average them $E(LOR_i)$

Shuster’s and SA estimates provide different results
We consider three different Bayesian hierarchical meta-analysis models:

1. Logistic regression
2. Arm-based approach
3. Adopting a beta hyperprior

Also consider constant (CTE) and heterogeneous (HTE) treatment effects with Bayesian models:

- For logistic regression and beta hyperprior models, we consider CTE and HTE models.
- We assume only HTE for arm-based models.
Bayesian Logistic Regression

Assume the following for binomial parameter $p_{ik}$ for study $i$ trt $k$

- **CTE model:** \[ \text{logit}(p_{ik}) = \mu_i + d \times I(k = 2) \]
- **HTE model:** \[ \text{logit}(p_{ik}) = \mu_i + \delta_i \times I(k = 2) \]

**CTE**
- $\mu_i \sim N(0, 100)$: baseline log odds for study $i$’s control group
- $d \sim N(0, 100)$: Assumed common log odds ratio

**HTE: Distinguish types of heterogeneity**
- **HTE-Logit** assumes
  - $\delta_i \mid d, \tau \sim N(d, \tau^2)$, $d \sim N(0, 100)$, $\tau \sim \text{Unif}(0, 2)$
  - $\tau$ measures the heterogeneity of LOR across studies
- **HTE-LogitSh** adds shrinkage prior for baseline effect $\mu_i \sim N(m_\mu, \tau_\mu^2)$
  - $m_\mu \sim N(0, 100)$ and $\tau_\mu \sim \text{Unif}(0, 2)$
Bayesian Arm-Based Model

- Model each arm’s risk within each study separately
  - Based on Hong et al. (2016) arm-based parameterization in network meta-analysis
- We simplify this model to fit an arm-based meta-analysis

\[ \text{logit}(p_{ik}) = \theta_k + \eta_{ik} \]

- \( \theta_k \) = log odds in arm \( k = 1, 2 \)
- \( \eta_{ik} \) = random to allow trt-specific risk heterogeneity across studies
- We assume \( (\eta_{i1}, \eta_{i2})^T \sim N_2((0, 0)^T, \Sigma) \)
- We use priors \( \theta_k \sim N(0, 10^2) \) and \( \Sigma^{-1} \sim \text{Wishart}(\Omega, 2) \)
  - \( \Omega = 0.02 \times I_2 \)
- Log odds ratio can be obtained using \( \theta_2 - \theta_1 \)
Bayesian CTE with Beta Priors

- Consider beta prior distributions
- For CTE, likelihood is rewritten as
  \[ y_{i1} \mid p_1 \sim Bin(n_{i1}, p_1) \text{ and } y_{i2} \mid p_2 \sim Bin(n_{i2}, p_2) \]
- Priors, such as
  \[ p_1 \sim Beta(\alpha, \beta) \text{ and } p_2 \sim Beta(\alpha, \beta) \]
  \[ \alpha \text{ and } \beta \text{ are prespecified.} \]
  \[ \text{We used } \alpha = \beta = 1 \]
Bayesian: HTE with Beta Hyperprior

- For HTE, likelihood is
  \[ y_{i1} \mid p_{i1} \sim \text{Bin}(n_{i1}, p_{i1}) \text{ and } y_{i2} \mid p_{i2} \sim \text{Bin}(n_{i2}, p_{i2}) \]

- Consider beta priors but different parameterization
  \[ p_{i1} \sim \text{Beta}(U_1 V_1, (1 - U_1) V_1) \text{ and } p_{i2} \sim \text{Beta}(U_2 V_2, (1 - U_2) V_2) \]

  Note: \( E[p_{ik}] = U_k \) and \( \text{var}[p_{ik}] = U_k (1 - U_k) / (V_k + 1) \)

- We assign Beta(1,1) hyperpriors to \( U_k \) and vague gamma hyperpriors to \( V_k, k = 1, 2 \)

If \( X \sim \text{Beta}(\mu \phi, (1 - \mu) \phi) \), \( E[X \mid \mu, \phi] = \mu \) and \( \text{var}[X \mid \mu, \phi] = \frac{\mu(1-\mu)}{\phi + 1} \)

In terms of \( \text{Beta}(\alpha, \beta) \) parameterization, \( \mu = \frac{\alpha}{\alpha + \beta} \) and \( \phi = \alpha + \beta \)
Simulation Studies of 15 Models: 9 Freq & 6 Bayesian

- Compare (1) bias, (2) MSE, & (3) coverage prob of LOR estimates
- Each simulated meta-analysis dataset has 30 studies
- Sample sizes: Control group’s $n_1 \sim Unif(50, 1000)$ & $n_2 = n_1$
- Risk for trt $k = 1, 2$ in study $i$:
  - $p_{ik} \sim Unif(p_k(1 - 0.5D), p_k(1 + 0.5D))$
  - $p_k$ depends on scenario
  - $D$ controls variability of risks: $D = 0, 1, 2$

- We consider the following risk parameters, $p_k$

<table>
<thead>
<tr>
<th>Null</th>
<th>Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>$p_2$</td>
</tr>
<tr>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

- We consider 18 scenarios, 10,000 simulated meta-analyses per scenario
Methods We Compare

Frequentist
- CTE
  - Naïve [Naive]
  - Peto [Peto]
  - Mantel-Haenszel (MH) [MH]
  - MH-Data Modification (DM) [MH(DM)]
  - Inverse Variance-DM [CTE-IV(DM)]
  - SGS-Unweighted [SGS-Unwgt]
  - SGS-Weighted [SGS-Wgt]
- HTE
  - DerSimonian & Laird IV-DM [HTE-DL(DM)]
  - Simple Average with $\tau_{DL}$ [SA(DM)]

Bayesian
- CTE
  - Logistic Regression [CTE-Logit]
  - Beta Hyperprior [CTE-Beta]
- HTE
  - Logistic Regression [HTE-Logit]
  - Logistic w/ Shrinkage prior [HTE-LogitSh]
  - Arm-based model [AB-Logit]
  - Beta Hyperprior [HTE-Beta]
Results for Log Odds Ratio: Null Case - Bias

D = 0

D = 2

Bias

-0.25 -0.20 -0.15 -0.10 -0.05 0.00 -0.25 -0.20 -0.15 -0.10 -0.05 0.00

Shapes:
- CTE model
- HTE model

Colors:
- Frequentist
- Bayesian

$D = 0$

$D = 2$

$p = 0.002$

$p = 0.005$

$p = 0.002$

$p = 0.005$

$p = 0.005$
Results for Log Odds Ratio: Null Case - MSE

D = 0

D = 2

Shape
- • CTE model
- ▲ HTE model

Color
- □ Frequentist
- △ Bayesian

MSE

HTE−Beta
AB−Logit
HTE−LogitSh
HTE−Logit
CTE−Beta
CTE−Logit
SA(DM)
HTE−DL(DM)
SGS−Wgt
SGS−Unwgt
CTE−IV(DM)
MH(DM)
MH
Peto
Naive
Results for Log Odds Ratio: Null Case - Coverage Prob

- Naive
- Peto
- MH
- MH(DM)
- CTE–IV(DM)
- SGS–Unwgt
- SGS–Wgt
- HTE–DL(DM)
- SA(DM)
- CTE–Logit
- CTE–Beta
- HTE–Logit
- HTE–LogitSh
- AB–Logit
- HTE–Beta

Shape
- ● CTE model
- ▲ HTE model

Color
- ● Frequentist
- ▲ Bayesian

Coverage Probability

D = 0
- p = 0.002
- p = 0.005

D = 2
- p = 0.002
- p = 0.005

0.4 0.6 0.8 1.0
0.4 0.6 0.8 1.0

HTE–Beta
AB–Logit
HTE–LogitSh
SA(DM)
HTE–DL(DM)
SGS–Wgt
SGS–Unwgt
HTE–Logit
CTE–Beta
CTE–Logit
SA(DM)
CTE–IV(DM)
MH(DM)
MH
Peto
Naive
Simulation Results: Summary for Null Case

- All frequentist estimators (which are moment-based estimators) and three Bayesian approaches modeling trt-specific risks (CTE-Beta, AB-Logit, and HTE-Beta) showed little to no bias
  - Bayesian models with priors on study-specific effects did OK
- CTE-Logit, HTE-Logit, and HTE-LogitSh showed relatively large biases
  - Data-generating mechanism and model structure do not agree
- MSE decreased as risks increased (i.e. more events)
- When between-study variability existed (D=2), SGS-Unwgt, HTE-Logit, HTE-LogitSh, and AB-Logit tended to have large MSEs
- D=2 case: CTE models (except SGS estimators) had poor coverage. Got worse as risks grew larger
Results for Log Odds Ratio: Alt Case - Bias

D = 0

D = 2

\( p = 0.002 \), \( p = 0.004 \)

\( p = 0.005 \), \( p = 0.01 \)

Shape
- ● CTE model
- ▲ HTE model

Color
- ● Frequentist
- ▲ Bayesian

- Naive
- Peto
- MH
- MH(DM)
- CTE–IV(DM)
- SGS–Unwgt
- SGS–Wgt
- HTE–DL(DM)
- SA(DM)
- CTE–Logit
- CTE–Beta
- HTE–Logit
- HTE–LogitSh
- AB–Logit
- HTE–Beta
Results for Log Odds Ratio: Alt Case - MSE

- Naive
- Peto
- MH
- MH(DM)
- CTE–IV(DM)
- SGS–Unwgt
- SGS–Wgt
- HTE–DL(DM)
- SA(DM)
- CTE–Logit
- CTE–Beta
- HTE–Logit
- HTE–LogitSh
- AB–Logit
- HTE–Beta

**Shape**
- CTE model
- HTE model

**Color**
- Frequentist
- Bayesian
Results for Log Odds Ratio: Alt Case - Coverage Prob

- **D = 0**
  - Naive
  - Peto
  - MH
  - MH(DM)
  - CTE–IV(DM)
  - SGS–Unwgt
  - SGS–Wgt
  - HTE–DL(DM)
  - SA(DM)
  - CTE–Logit
  - CTE–Beta
  - HTE–Logit
  - HTE–LogitSh
  - AB–Logit
  - HTE–Beta

- **D = 2**
  - Naive
  - Peto
  - MH
  - MH(DM)
  - CTE–IV(DM)
  - SGS–Unwgt
  - SGS–Wgt
  - HTE–DL(DM)
  - SA(DM)
  - CTE–Logit
  - CTE–Beta
  - HTE–Logit
  - HTE–LogitSh
  - AB–Logit
  - HTE–Beta

**Shape**
- CTE model: ●
- HTE model: ▲

**Color**
- Frequentist: ○
- Bayesian: ▲

**Coverage Probability**
- Coverage Probabilities for different methods under different conditions.
• Naïve, MH, SGS, and HTE-Logit had small biases with D=0 and D=2
• SGS-Unwgt had large MSE in lowest risk scenario
  \( (p_1 = 0.002, p_2 = 0.004) \)
• When D=2, SGS, HTE-Logit, and HTE-Beta gave coverage probability close to nominal level (0.95) across all scenarios
Nissen & Wolski (NW) Data: Description of Studies

- **MI:**
  - 159 events in 19,509 among pts in rosiglitazone-containing arms
    - Pooled risk = 0.00815
  - 136 events among 16,022 pts in arms w/o rosiglitazone
    - Pooled risk = 0.00849
  - Naïve pooled estimate of LOR: \(-0.041\) (95% interval: -0.271, 0.189)

- **CV death:**
  - 105 events among 19,509 pts in rosiglitazone-containing arms
    - Pooled risk = 0.00538
  - 100 events among 16,022 pts in arms w/o rosiglitazone
    - Pooled risk = 0.00624
  - Naïve pooled estimate of LOR: \(-0.149\) (95% interval: -0.424, 0.126)
## Summary of Data: Studies with Zeros

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. of studies</th>
<th>0-0 studies</th>
<th>0 in control</th>
<th>0 in trt</th>
<th>Non-zero studies</th>
<th>Corr ($pc_i$, $nc_i$)</th>
<th>Corr ($pa_i$, $na_i$)</th>
<th>$E(pc_i)$</th>
<th>$E(pa_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data1</td>
<td>MI</td>
<td>56</td>
<td>15</td>
<td>35</td>
<td>21</td>
<td>0.312</td>
<td>0.181</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>Data2</td>
<td>MI</td>
<td>55</td>
<td>15</td>
<td>35</td>
<td>21</td>
<td>0.154</td>
<td>-0.003</td>
<td>0.004</td>
<td>0.006</td>
</tr>
<tr>
<td>Data3</td>
<td>MI</td>
<td>42</td>
<td>9</td>
<td>29</td>
<td>13</td>
<td>-0.024</td>
<td>-0.156</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>Data4</td>
<td>MI</td>
<td>13</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>0.243</td>
<td>-0.147</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td>Data1</td>
<td>CV</td>
<td>56</td>
<td>29</td>
<td>44</td>
<td>33</td>
<td>0.318</td>
<td>0.202</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>Data2</td>
<td>CV</td>
<td>55</td>
<td>29</td>
<td>44</td>
<td>33</td>
<td>0.042</td>
<td>-0.030</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>Data3</td>
<td>CV</td>
<td>42</td>
<td>20</td>
<td>35</td>
<td>23</td>
<td>0.035</td>
<td>-0.063</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td>Data4</td>
<td>CV</td>
<td>13</td>
<td>4</td>
<td>10</td>
<td>6</td>
<td>0.563</td>
<td>0.104</td>
<td>0.001</td>
<td>0.002</td>
</tr>
</tbody>
</table>
NW Data Analysis: Log Odds Ratio - MI

All studies (56 studies)
Outcome: MI

All studies except RECORD trial (55 studies)
Outcome: MI

Trials comparing rosiglitazone+X vs. X alone (42 studies)
Outcome: MI

Trials comparing rosiglitazone vs. placebo (13 studies)
Outcome: MI
NW Data Analysis: Risk Difference - MI

All studies (56 studies)
Outcome: MI

All studies except RECORD trial (55 studies)
Outcome: MI

Trials comparing rosiglitazone+X vs. X alone (42 studies)
Outcome: MI

Trials comparing rosiglitazone vs. placebo (13 studies)
Outcome: MI
NW Data Analysis: Risk Difference - CV death

All studies (56 studies)
Outcome: CV death

- Naive
- MH
- MH(DM)
- CTE–IV(DM)
- SGS–Unwgt
- SGS–Wgt
- HTE–DL(DM)
- CTE–Logit
- CTE–Beta
- HTE–Logit
- HTE–Log
- HTE–LogitSh
- HTE–LogSh
- AB–Logit
- AB–Log
- HTE–Beta

Risk difference: -0.008 to 0.004

All studies except RECORD trial (55 studies)
Outcome: CV death

- Naive
- MH
- MH(DM)
- CTE–IV(DM)
- SGS–Unwgt
- SGS–Wgt
- HTE–DL(DM)
- CTE–Logit
- CTE–Beta
- HTE–Logit
- HTE–Log
- HTE–LogitSh
- HTE–LogSh
- AB–Logit
- AB–Log
- HTE–Beta

Risk difference: -0.008 to 0.004

Trials comparing rosiglitazone+X vs. X alone (42 studies)
Outcome: CV death

- Naive
- MH
- MH(DM)
- CTE–IV(DM)
- SGS–Unwgt
- SGS–Wgt
- HTE–DL(DM)
- CTE–Logit
- CTE–Beta
- HTE–Logit
- HTE–Log
- HTE–LogitSh
- HTE–LogSh
- AB–Logit
- AB–Log
- HTE–Beta

Risk difference: -0.008 to 0.004

Trials comparing rosiglitazone vs. placebo (13 studies)
Outcome: CV death

- Naive
- MH
- MH(DM)
- CTE–IV(DM)
- SGS–Unwgt
- SGS–Wgt
- HTE–DL(DM)
- CTE–Logit
- CTE–Beta
- HTE–Logit
- HTE–Log
- HTE–LogitSh
- HTE–LogSh
- AB–Logit
- AB–Log
- HTE–Beta

Risk difference: -0.008 to 0.004
Naive and CTE-Beta provided similar results
- They ignore heterogeneity
CTE-IV(DM) and HTE-DL(DM) provided exactly the same results because $\hat{\tau} = 0$ in HTE-DL(DM)
When all studies were included, posterior means of $\tau$ in HTE-Logit and HTE-LogitSh are 0.12 and 0.14.
SGS-Wgt estimates always provided much narrower 95% confidence intervals than SGS-Unwgt
- This trend was not observed in our simulations. It may be due to correlation between study design (sample size) and risk induced in the NW data.
SGS-Unwgt very different from SGS-Wgt when only look at placebo-controlled RCTs
SGS-Unwgt and SA provided very different estimates.
Conclusion

- The best way to estimate treatment effects correctly with rare events is to conduct large RCTs (and combine information from them).
- Simulation results suggest that one interpret associations carefully in rare (very low frequency) event settings.
- When the outcome is really uncommon, the normality random effect assumption might not hold.
- Effect estimates vary when applying different models; inference can differ, too.
- Future work, we will relax the normality assumption. In addition, we will consider nonparametric approaches.
References: