Meta-analysis in randomized clinical trials with low event rates

Jonathan J. Shuster, Ph.D
(jshuster@biostat.ufl.edu)
Research Professor of Biostatistics
Univ. of Florida, College of Medicine

Expectations?

Outline of Talk

- **What is special about Low Event Rate Meta Analysis?**
- **Fixed vs. Random Effects: Analytic Models**
- **Zero event treatment arms in Fixed Effects vs. Random Effects.**
- **Argument on weighting of trials**

Outline (Continued)

- **The Rosiglitazone Example**
- **Issues of Jeopardy for Drug/Device Companies in Meta Analysis**
- **Data and Safety monitoring of products**
- **Conclusions and recommendations**

What is special about low event rates

- These low event rate meta analyses rely on a large number of studies, each with low event rates. The asymptotic behavior is not within an individual study, but due to the large collection of studies. Individual estimates within studies will be poorly behaved.

Fixed Effects Binomial Approach

- The Gold Standard for fixed effects 2-sample binomial meta analysis is via a presumed common odds ratio across studies (the fixed effect). That odds ratio is estimated by Maximum Likelihood (equivalent to Logistic Regression with categorical covariates).
- Another fixed effect model could assume a common difference in proportions, but few do things this way.

Typical individual Study j Data may look like this

Est OR= $F_{j2} S_{j1} / F_{j1} S_{j2}$.	Arm 1	Arm2
Success	S_{j1}	S_{j2}
Fail	F_{j1} (can be zero)	F_{j2} (can be zero)

7

Fixed Effects Approach

- The contribution to the likelihood of a study is proportional to
- $P_{j1}^{F_{j1}} Q_{j1}^{S_{j1}} P_{j2}^{F_{j2}} Q_{j2}^{S_{j2}}$ **(1)**
- **Where the P's are the failure rates, Q's are the success rates (1-P)'s, and the population odds ratios, $P_{j2} Q_{j1} = \Psi P_{j1} Q_{j2}$ with Ψ constant over studies.**
- **If $F_{j1} = F_{j2} = 0$, the Maximum value of (1) ($Q_{jk} \rightarrow 1$ above)=1 (no contribution) (Can exclude study) (Ψ plays no role)**

8

Fixed Effects Approach (Continued)

- $P_{j1}^{F_{j1}} Q_{j1}^{S_{j1}} P_{j2}^{F_{j2}} Q_{j2}^{S_{j2}}$ **(1)**
- **If only one of the F's is zero (say F_{j1}), the constraint that there is a common odds ratio $\Psi = P_{j2} Q_{j1} / P_{j1} Q_{j2}$ across studies (1) becomes:**
- $\Psi^{S_{j1}} P_{j1}^{S_{j1}} Q_{j2}^{S_{j1}+S_{j2}} P_{j2}^{F_{j2}-S_{j1}}$ **(2)**
- **This forces a non zero maximum for P_{j1}**

9

Random Effects

- **Studies** Represent a random sample taken from a hypothetical target population.
- Goal is to make an inference about "the effect size" in the target population from the effect sizes in the sample.
- For binomial data, even using odds ratios (relative risk for low events), there are choices.

10

Choices (Ignore rare event status for now)

- Average odds Ratio (weighted or unweighted)
- Take an average of the natural logs of the odds ratios (weighted or unweighted)
- Take the ratio of the average failure rate in Arm 1 to the Average failure rate in Arm 2. (or log of this ratio)

11

Average odds Ratio (weighted or unweighted)

- What are we estimating:
- Suppose we had odds ratios of 4.0, 1.0, and 0.25 with equal probabilities in our universe. The average odds ratio would be 5.25/3=1.75 (Arm 2 to Arm 1)
- Reverse the roles of Arm 2 and Arm 1:
- 4.0→0.25, 1.0 →1.0, and 0.25→4.0. So the average odds ratio of Arm 1 to Arm 2 is also 1.75. Since 1.0 is the dividing line for equality, this approach is clearly unacceptable.

12

Take an average of the natural logs of the odds ratios?

- When you use logs, and reverse the treatments, you get an estimate in the log scale as the negative of that in the original scale, and if you take antilogs you get 1/old estimate. (You get log(OR)=0 in both cases in our example).

13

Issue of No Events on one Arm

- The real issue is zeros on either or both arms. If on both, the estimate is undefined, and if on one, you get +/- ∞ (Don't even get into continuity corrections, as they present huge bias.)

14

Ratio of the average failure rate Arm1: Arm 2

- Strategy: Treat a patient on a randomly selected study on Arm 1; Independently treat a second patient on a randomly selected study on Arm 2.
- What is the ratio of failure probabilities for the Arm 1 patient to the Arm 2 patient (a different but reasonable relative risk definition).

15

Reciprocal Consistency

- Note that you are estimating a ratio of means, and so if you estimate 2:1 and I estimate 1:2, we get the reciprocals of each other as point estimates.
- This is our chosen outcome measure for low event rate random effects meta analysis (We will actually work with the log of this but then take antilogs).

16

Unweighted vs. Weighted Estimation

Suppose we wish to estimate the population mean of C-Reactive Protein Levels in Lupus Patients. We have a representative sample of subjects with 3-20 repeated measures. Do we?

17

Do we

- Estimate the mean by the average of all of the measurements?
- Estimate the mean by taking the mean of the individual subjects and average these?
- Estimate the mean by taking the mean of each individual subject and taking a weighted mean of these, inversely proportional to the square of the estimated within subject standard error?

18

Argument for Equal Weighting of Trials

Standard Sampling Model

- We have a target population of Studies from which we draw a random sample of studies from the population
- We want to estimate the Mean "effect size" in the Population of Studies (not the population of people).

19

The Universal Binomial Model

(1) $\mu_{jk} = \boldsymbol{\theta}_k + \delta_{jk}$

(2) $P_{jk} = \mu_{jk} + \varepsilon_{jk} = \boldsymbol{\theta}_k + \delta_{jk} + \varepsilon_{jk}$

(3) Draw a study j at random from a population of studies, and the true target population event rates for study j are μ_{j1} and μ_{j2} for treatments k=1 and 2. The observed proportion is then obtained P_{jk} for each study sampled, and each treatment. θ_1 and θ_2 are the target quantities of interest. The δ and ε are random errors with means of zero.

20

What weighted advocates forget

- Note that $E(P_{jk}) = \theta_k$ under equal weighting.
- For unequal weighting, the weights are random variables determined by the randomly selected design. If these weights are associated with outcome, then the estimates are biased. A simple example will follow next.

21

Meta Analysis of Sudafed (NAR at 60 Min)

Study	N/Group	Mean	SE	Weight
1	16	-5.63	0.74	17%
2	10	-3.88	0.36	**71%**
3	16	-3.55	2.15	1.9%
4	15	-1.85	1.59	3.6%
5	15	-0.08	1.74	3.0%
6	16	1.35	2.11	2.0%
7	25	1.59	2.30	**1.7%**

22

Meta analysis of Sudafed

- Random effects: Equal weight Est Effect size=-1.72 (SE=1.05) (P=0.15)
- Weights per last column: Est Effect size -3.79 (SE=0.79) (P<0.001)
- DerSimonian and Laird method. Meta analysis in Clinical Trials. Controlled Clinical Trials 7:177-188 (1986)

23

Reference

- **Meta-Analysis of the Efficacy of a Single Dose of Phenylephrine 10 mg Compared with Placebo in Adults with Acute Nasal Congestion Due to the Common Cold**
- **Kollar et. al. Clinical Therapeutics/Volume 29, Number 6, 2007, 1057-1070.**

24

"Large Sample" Estimation

- $P_{jk} = F_{jk} / N_{jk}$.
- Failures divided by sample size for Study j, Treatment k. By our model
- (P_{j1}, P_{j2}) are iid vectors with mean (θ_1,θ_2)
- The relative risk $RR= \theta_1/\theta_2$ is estimated by $RR^*=\Sigma P_{j1} / \Sigma P_{j2})$

25

More on Large Sample Estimation

- From the Delta Method for large samples $\log(RR^*)=\log(\Sigma P_{j1})-\log(\Sigma P_{j2})$ is asymptotically normal with mean log(RR) and variance:

$SE^2= \{(\sigma_1/ \theta_1)^2 + (\sigma_2/ \theta_2)^2 -2\rho(\sigma_1 \sigma_2 /\theta_1\theta_2)\}/M$

26

More on Large Sample Estimation

M=number of studies, σ_k the population Standard deviations of the P_{jk} (k=1 and k=2) and ρ the population correlation coefficient between P_{j1} and P_{j2} .

To obtain a consistent estimator of the variance of log(RR*), SE^2, we replace the population means, standard deviations and correlations by the sampling values for P_{j1} and P_{j2}

Log(RR*)+/1.96SE forms a large sample 95% CI for Log(RR).

27

Nissen-Wolski Study

Inclusion Criteria for Trials

- Phase 2/3/4
- 24+ weeks of Rosiglitazone on one arm
- Randomized to two treatments, one with Rosiglitazone one without.

28

Description of Trials

- 116 Screened, 48 Eligible
- Endpoints: Myocardial Infarction (MI) and Cardiac Death (CD)
- 6 Trials were excluded for no events (MI or CD)
- About 30,000 patients are included from the 42 trials
- 38 Trials had at least one MI, 23 Trials had at least one CD

29

Studies Description

- 4 Trials were for conditions other than Type II Diabetes ((Alzheimer's Disease, Impaired glucose tolerance, and Psoriasis {2})

30

Studies Description

- Had widely differing doses, differing follow-up, differing control medications, differing eligibility, and differing concomitant medications

31

Nissen-Wolski Data

Study (Time)	N_{1j}	$F_{1j\,(MI)}$	$F_{1j\,(CD)}$	N_{2j}	$F_{2j\,(MI)}$	$F_{2j\,(CD)}$
1	774	1	0	185	1	0
2	583	2	0	142	0	0
3	442	1	1	112	0	0
4	394	1	1	124	0	0
5	1172	1	1	377	0	0
6	706	0	1	325	0	0
7	284	1	0	135	0	0
8	196	0	1	96	0	0
Etc						
Etc						
39	231	1	1	242	0	0
40	254	1	0	272	0	0
41	43	0	0	47	1	0
42	1456	27	2	2895	41	5

32

Statistical Analysis: Nissen

- Step 1: Test hypothesis that there is a Common Odds Ratio by Cochran Q Test for the included studies (38 for MI and 23 for CD)
- Step 2: If non-significant, apply fixed effects, essentially logistic regression. If significant, (P<0.10) presumably they run some unnamed random effects meta-analysis.
- P for Cochran Q was not stated in paper, but we got P=0.78 (MI) and P=0.99 (CD), by STATXACT

33

Statistical Analysis Random Effects

- RR-est=ΣP_{j2} / ΣP_{j1} (Ratio Estimate) and SE via the Delta method per earlier
- Shuster, Jones, Salmon: Fixed vs. random effects meta-analysis in rare event studies: The Rosiglitazone link with myocardial infarction and cardiac death. *Statistics in Medicine*, 2007 Oct; 26: 4375-4385.
- Zeros are no problem in this context

34

Myocardial infarction Results

Method	Estimated RR	95% CL	P-Value Two-sided
Random Effects	1.51	0.91-2.48	0.11
Fixed Effects			
Nissen and Wolski	1.43	1.03-1.98	0.03

35

Cardiac Death Results

Method	Estimated Relative Risk	95% CL	P-Value Two-sided
Random Effects	2.37	1.38-4.07	0.0017
Fixed Effects			
Nissen and Wolski	1.64	0.98-2.74	0.06

36

Cardiac Death Issue by Study

Rosi Event?→ Control Event?↓	No	Yes
No	25	**15**
Yes	**2**	6

37

Jeopardy

Industry Perspectives
The Public is Scrutinizing
It is haphazard

38

Jeopardy

Dr. Nissen is watching for Cardiac Endpoints
Dr. Sneezer is watching for Allergic Reactions

39

Jeopardy

Dr. Renaldo is watching for Kidney problems
Dr. Hepato is watching for adverse liver events
Dr. McBrain is watching for CNS events

40

Time Jeopardy

- **They are all watching serially over time**

41

What Gets Published?

- Negative Findings?
- Are serial looks disclosed?

42

Impact of Meta Analysis on Ongoing Rosiglitazone Trials

- Two large trials that would have settled the issue of MI were largely scuttled by the Nissen-Wolski publication (RECORD and BARI-2D).
- Rosiglitazone is one of several agents designed to eliminate or delay the need for regular insulin shots.

43

Data and Safety Monitoring of Products

- We all accept these committees for individual trials
- Why not have these committees for products, from first human testing until going off patent?
- Side effects are multivariate problems and time-series problems at the same time.

44

Make-up of Committee

- No member of the company should on the committee
- Make-up should at a minimum include Specialists in the subject area (MDs, nurses, PhDs as appropriate), an ethicist, a biostatistician (expert in sequential methods and multivariate methods), a lay person.

45

Meeting Procedures (at least once a year)

- Open session where drug/device company can participate
- Closed session to determine if action is needed
- Committee is charged with conducting its own periodic meta-analysis and reviewing any external meta analyses.

46

Conclusions and Recommendations

- **Fixed vs. Random Effects: Fixed effects should not be used unless the studies are pure replications with the same eligibility.**
- Deciding fixed vs. random effects via a Diagnostic test is invalid statistically. These tests have poor sensitivity and specificity.

47

Conclusions and Recommendations

- What about the test for Relative Risk=1 by fixed effects?
- This validly tests the null hypothesis that all individual tests have RR=1, but it is too narrow a null hypothesis to be useful, unless fixed effects are considered valid from the start.

48

Conclusions and Recommendations

- For Random Effects Meta Analysis, avoid bias by weighting all studies equally.
- Treat the (P_{1j}, P_{2j}) as independent identically distributed vectors of observed proportions from the target population.

49

Conclusions and Recommendations

- Don't over interpret Meta Analysis
- Recognize potential selection bias in terms of what drug/device is studied, what endpoint is studied (why chosen), and at what timing was it studied.

50

Conclusions and Recommendations

- Independent Data and Safety Monitoring Committee for Products is Recommended for Drug/Device companies from first human study to end of patent.

51

Expectations Realized?

52

Thank You

53

Actual Dilbert Solutions

- Invest $100 in a tax-free no load bond fund at 5%, and you will have $1,000,000 in August 2197 (188.8 years)
- Invest $100 in taxable CDs at 5% (33.33% bracket), and you will have $1,000,000 in September 2289 (280.9 years)

54