Meta-analysis of clinical trials, particularly of rare adverse events

Peter Lane

Research Statistics Unit

Outline

1. What is meta-analysis?
2. How does it work?
3. Efficacy of SSRIs:
4. Safety of Avandia:
5. Trials with no events
6. Graphical summary

What can go wrong?

Acknowledgements:

- James Roger & Valerii Fedorov (RSU), for many discussions
- George Quartey (now at Roche) with whom I've developed a two-day course on meta-analysis

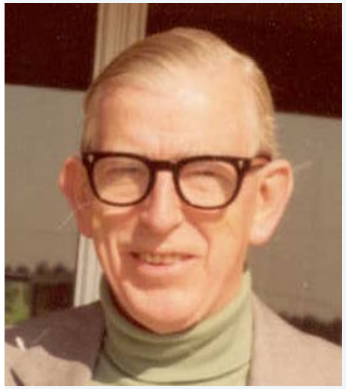
1. What is meta-analysis?

The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings.

(Glass, 1978)

History

- Fisher (1940s) started the ball rolling
  - Agricultural trials
  - Combining p-values



- Cochran (1950s) laid down the basics:
  - Inverse-variance weighting
  - Testing homogeneity

- Glass (1970s) coined the name
  - Showed efficacy of psychotherapy

Early use in Medicine

- Yusuf et al (1985) used a meta-analysis to show that long-term beta-blockade following discharge from the coronary care unit after MI reduced mortality
- The Early Breast Cancer Trialists' Collaborative group (1988) showed that tamoxifen reduced mortality in women over 50 with early breast cancer

Cochrane collaboration

- Started in 1993
- Towards evidence-based medicine
- Over 4,000 systematic reviews
  - Available on-line:

    http://www.cochrane.org/
  - Produced by volunteer healthcare professionals, overseen by editorial teams
  - Most systematic reviews include a meta-analysis (term usually used to refer to the phase of combining results)

What data?

- Individual patient data
  - Available for in-house meta-analysis
  - Allows analysis of covariates
  - Methodology as for multi-centre trials
- Summary data
  - Use estimate and s.e. from each study
  - Range of special methods

2. How does it work?

- Summary data: “inverse variance” method is commonest
  - Statistic (e.g. treatment difference) from each study
  - Standard error of each statistic
- Weight each estimate by inverse variance
  - i.e. $1/s.e.^2$
  - Imprecise studies make less contribution
  - Studies contribute in proportion to the number of patients (if variability is the same)
  - Studies contribute in inverse proportion to the variability (if number of patients is the same)

Heterogeneity

- Inverse-variance method provides a combined estimate and a standard error
  - confidence intervals and a p-value can be derived
- Heterogeneity of the estimates can be calculated
  - heterogeneity is central in the interpretation
  - used to help decide whether the studies should all be combined
- Clinical heterogeneity is at least as important as statistical heterogeneity

Example: dentifrices

Study	*N*1	Mean1	SD1	*N*2	Mean2	SD2
1	134	5.96	4.24	113	6.82	4.72
2	175	4.74	4.64	151	5.07	5.38
3	137	2.04	2.59	140	2.51	3.22
4	184	2.70	2.32	179	3.20	2.46
5	174	6.09	4.86	169	5.81	5.14
6	754	4.72	5.33	736	4.76	5.29
7	209	10.10	8.10	209	10.90	7.90
8	1151	2.82	3.05	1122	3.01	3.32
9	679	3.88	4.85	673	4.37	5.37

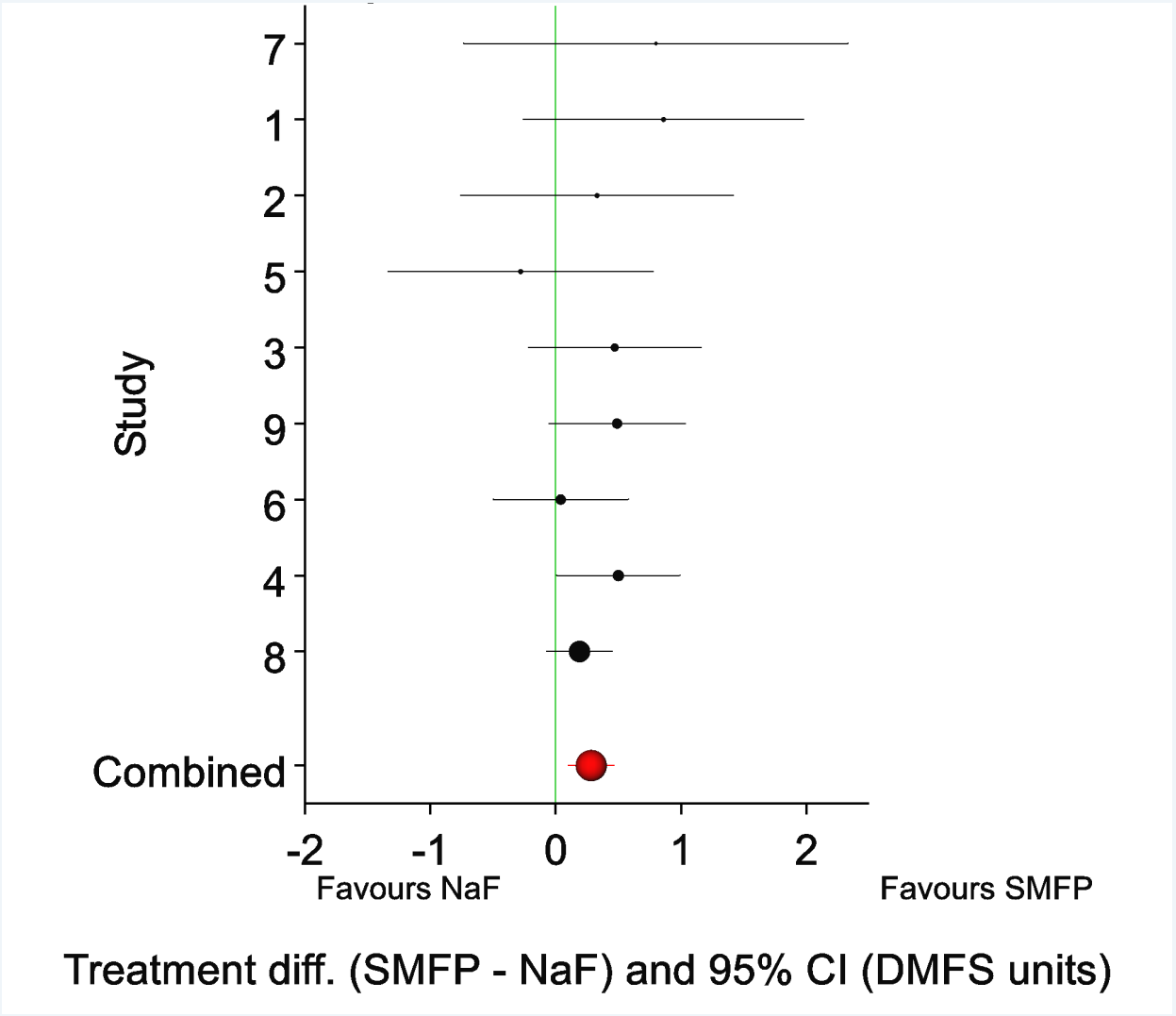
Weights

- SEs can then be calculated

$$SE = SD^{*}\sqrt{\{1/N_1 + 1/N_2\}}$$

Study	Diff	SE	Weight	$w = 1/SE^2$
1	0.86	0.57	3	
2	0.33	0.55	3	
3	0.47	0.35	8	
4	0.50	0.25	16	
5	-0.28	0.54	3	
6	0.04	0.28	13	
7	0.80	0.78	2	
8	0.19	0.13	56	
9	0.49	0.28	13	

Forest plot of NaF vs SMFP



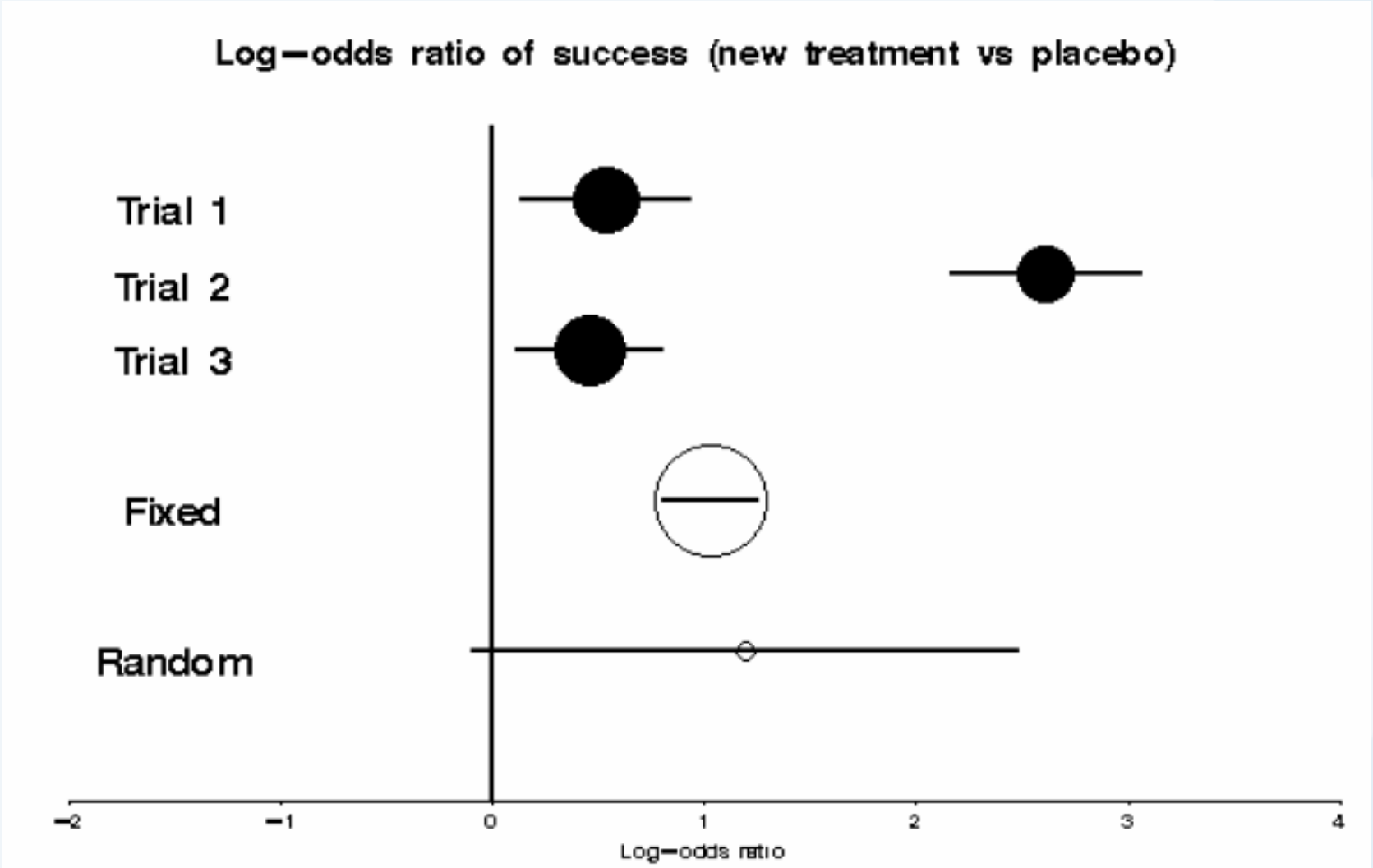
Random effects

- Fixed-effects method assumes same difference in all studies
- Random-effects method assumes a distribution
  - Add between-study variance to the model
- DerSimonian-Laird method adjusts inverse-variance formula:
  - $Q = \Sigma w(d–d^*)^2$ where d is diff., d^* is combined diff.
  - $\tau^2 = (Q–(k–1)) / (\Sigma w – \Sigma w^2/\Sigma w)$ or 0 (if negative)
  - $w^* = 1 / (SE^2 + \tau^2)$
- Cochran Q used as indicator of heterogeneity
  - Or $I^2 = 100(Q–(k–1))/Q$ or 0 (if negative)

Dentifrice heterogeneity

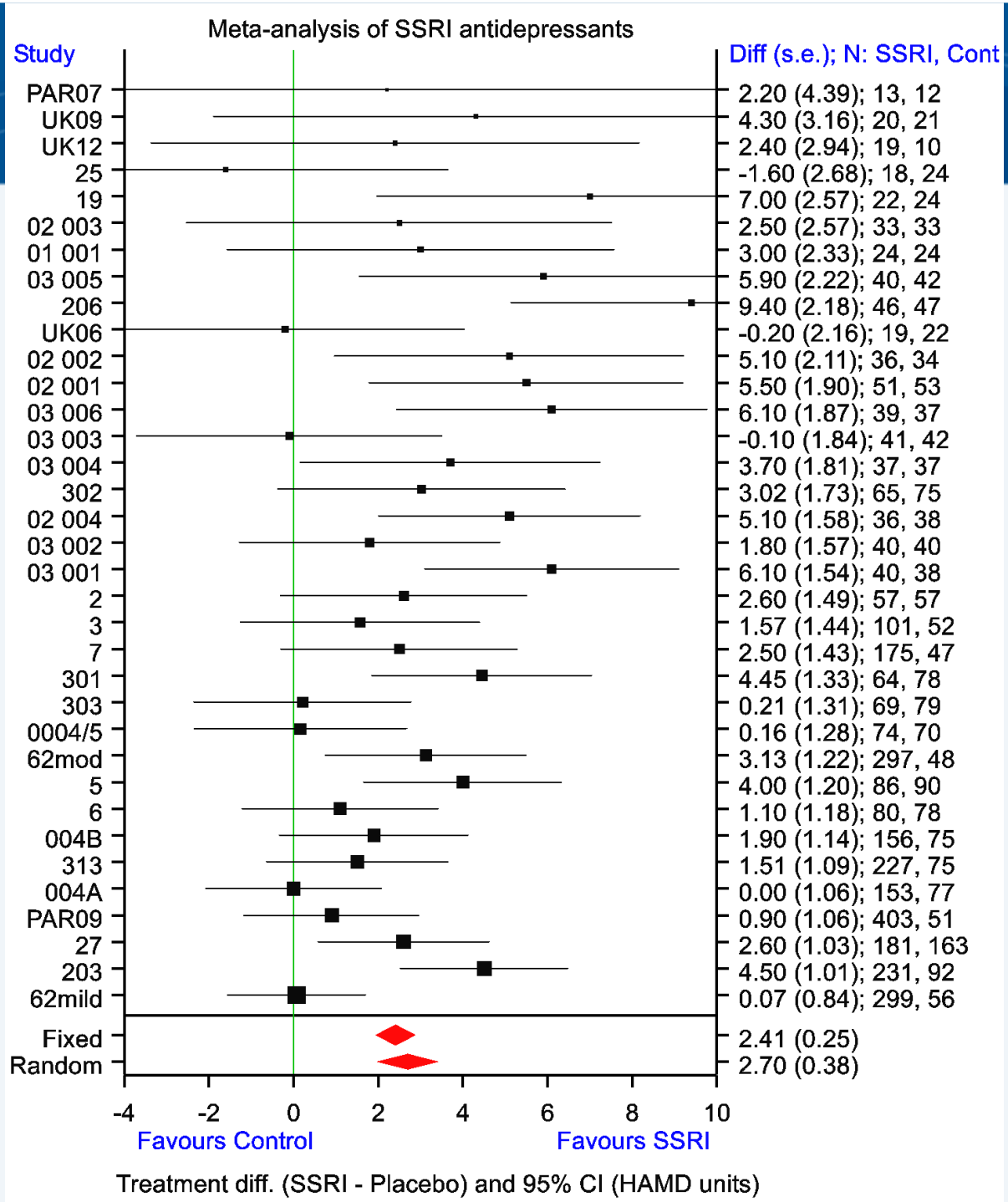
- Q = 5.4 (χ^2-statistic with df=8 if no heterogeneity)
- $\tau^2 = 0$
- $I^2 = 0$
- So the random-effects estimate is identical to the fixed-effect estimate

But what if we get this?



3. Efficacy of SSRIs

- Kirsch et al (PLoS Medicine, Feb 2008, and in Significance, 2008 Issue 2)
  - RCTs using SSRIs, assessing Hamilton depression scale
  - All trials provided in response to request to FDA
  - Paroxetine, fluoxetine, venlafaxine, nefazodone
  - 35 trials: 16 of paroxetine (cf. 352 trials on GSK's register!)
- Used naïve pooled summary, and naïve regression
  - No adjustment for imbalance of treatment arms
- Used naïve regression to investigate heterogeneity
  - Regressed mean change (from unpaired treatment arms) on baseline severity

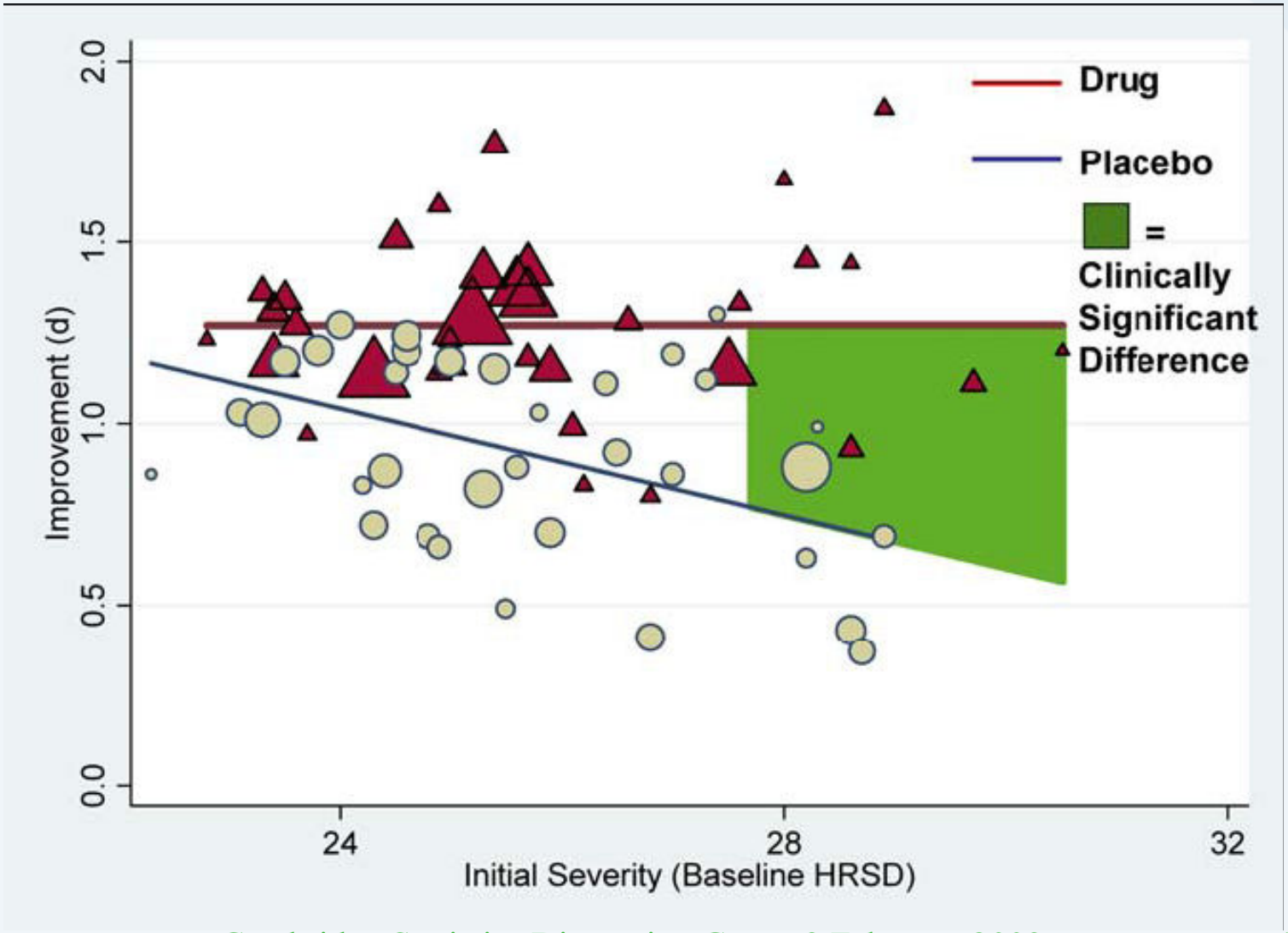

Meta-analysis of SSRI antidepressants
Study
Diff (s.e.); N: SSRI, Cont
PAR07 2.20 (4.39); 13, 12
UK09 4.30 (3.16); 20, 21
UK12 2.40 (2.94); 19, 10
25 -1.60 (2.68); 18, 24
19 7.00 (2.57); 22, 24
02 003 2.50 (2.57); 33, 33
01 001 3.00 (2.33); 24, 24
03 005 5.90 (2.22); 40, 42
206 9.40 (2.18); 46, 47
UK06 -0.20 (2.16); 19, 22
02 002 5.10 (2.11); 36, 34
02 001 5.50 (1.90); 51, 53
03 006 6.10 (1.87); 39, 37
03 003 -0.10 (1.84); 41, 42
03 004 3.70 (1.81); 37, 37
302 3.02 (1.73); 65, 75
02 004 5.10 (1.58); 36, 38
03 002 1.80 (1.57); 40, 40
03 001 6.10 (1.54); 40, 38
2 2.60 (1.49); 57, 57
3 1.57 (1.44); 101, 52
7 2.50 (1.43); 175, 47
301 4.45 (1.33); 64, 78
303 0.21 (1.31); 69, 79
0004/5 0.16 (1.28); 74, 70
62mod 3.13 (1.22); 297, 48
5 4.00 (1.20); 86, 90
6 1.10 (1.18); 80, 78
004B 1.90 (1.14); 156, 75
313 1.51 (1.09); 227, 75
004A 0.00 (1.06); 153, 77
PAR09 0.90 (1.06); 403, 51
27 2.60 (1.03); 181, 163
203 4.50 (1.01); 231, 92
62mild 0.07 (0.84); 299, 56
Fixed 2.41 (0.25)
Random 2.70 (0.38)
-4 -2 0 2 4 6 8 10
Favours Control
Favours SSRI
Treatment diff. (SSRI - Placebo) and 95% CI (HAMD units)

Combined estimate

- Naïve combined estimate: 1.80 units
- Fixed-effect method: 2.41 (s.e. 0.25)
- Random-effect method: 2.70 (s.e. 0.38)
- Significant heterogeneity (Q = 71.2, df=34; I^2 = 53%)

Drug	*Ntrial*	*Fixed*	*Random*
Fluoxetine	5	1.65 (0.55)	2.06 (1.07)
Nefazodone	8	1.63 (0.44)	1.65 (0.49)
Paroxetine	16	3.22 (0.47)	3.38 (0.61)
Venlafaxine	6	3.23 (0.53)	3.54 (1.06)

Regression analysis



4. Safety of Avandia

- Nissen & Wolski (2007)
  - RCTs using rosiglitazone, assess MI and CV death
  - Phase II to IV, ≥ 24 weeks, randomized comparator
  - 48 trials: six not detailed because there were no events
  - All 42 on the GSK Register, plus six candidates for those omitted
- Used the Peto method
  - Reasonable for rare events when effect is not large
  - Side-steps problems with trials that have events in only one arm (10 trials have no events, 20 have none under one treatment → only 18 trials with events in both groups)
  - Poor behaviour when highly unbalanced (8:1 or more) (Refs 3 & 5)
  - Biased estimates when the effect is substantial
  - Simulation with this set shows that the combined estimate has little bias and the properties of the significance test are satisfactory

48 trials

Study	Rosiglitazone			Comparator			Duration
	N	MI	CV	N	MI	CV	(weeks)
011	357	2	1	176	0	0	24
020	391	2	0	207	1	0	52
…							
dream	2635	15	12	2634	9	10	156
adopt	1456	27	2	2895	41	5	208
282	70	0	0	75	0	0	24
369	25	0	0	24	0	0	26
096	232	0	0	115	0	0	26
044	101	0	0	51	0	0	26
325	196	0	0	195	0	0	24
004	676	0	0	225	0	0	24

Controversial issues

1. Trials not designed to study CV problems:
   - Identification of events not planned and recorded as carefully as in studies designed for this purpose
2. Comparator groups vary widely:
   - Placebo, Metformin, Sulfonylurea, Insulin
3. Rosiglitazone as monotherapy or as adjunct therapy
4. Populations vary widely:
   - Mild diabetics, insulin-treated diabetics, psoriasis patients
5. Different doses of rosiglitazone: 2mg to 8mg
6. Duration of treatment varies widely: 24 to 208 weeks
7. Events were rare:
   - Less than 2% pa in all but five treatment groups
   - No incidence at all in nearly half the treatment groups

All of these cast doubt on the results and possible interpretation of the following meta-analyses

Naïve analysis

- Looking at the pooled results gives the impression that there is no treatment effect at all:
  - 0.56% incidence of MI under Comp and 0.51% under Rosi (i.e. about 5 patients per 1000 in the trials had a heart attack on average)
- This ignores the treatment imbalance (Simpson's Paradox)
  - ADOPT trial was large, 2:1 randomization (Comp:Rosi), and a higher incidence rate (four-year study)
  - Naturally more MIs reported under Comp in this trial, and the Treatment effect is confounded with the variation in overall incidence
- Moral: beware of pooling information without stratifying by Study

Scale of analysis

- Analysing risk difference has problems:
  - Risks are very likely to be related to duration
  - Scientifically, treatment likely to multiply risks rather than add to them
  - Wide range of incidence: 0.06% in Study 330 (psoriasis) to 3.1% in Study 211 (diabetics with CHF)
- Analysing relative risks has problems:
  - No information from 30 trials, if use log scale
  - But handles duration and is multiplicative
- Odds ratio very similar to relative risk
  - One-step method (Peto) excludes only 0-0 trials

Risk difference

- Inverse-variance method (excluding 0-0 trials):
  - 0.18% greater risk on Rosi, i.e. 2 per 1,000, NNH=556
  - 95% CI [0.07%, 0.28%], p=0.001
  - Weights: 40% '330', 9% Dream, 2% Adopt, 0% '0-0'

    0-0 trials would have infinite weight

    Psoriasis trial 330 has nearly half the weight
- Add 0.5 correction factor for trials with a zero
  - 0.02% [–0.12, 0.17] p=0.74
  - Weights: 13% '330', 17% Dream, 3% Adopt, 22% '0-0'
- Estimate depends entirely on chosen factor

Mantel-Haenszel method

- Originally developed for odds ratio, but extended for relative risk and risk difference
  - Recommended method by Cochrane Collaboration
- Study weight is $n_1 n_2 / N$
  - n_i is no. of patients with Treatment i, $N = n_1 + n_2$
  - So weight depends only on size of trial
- Mantel-Haenszel estimate:
  - 0.19% [0.01%,0.36%], p=0.034
  - Weights: 4% '330', 20% Dream, 14% Adopt, 12% '0-0'
  - Fortuitously close to the IV method without 0-0 trials, despite very different weighting

Adjust for duration

- Adjusting for duration of trial:
  - Risk range is unchanged (330 and 211 both 1-year), so still have problem of underlying model
  - Interpretation is improved in terms of exposure to drug
- Inverse-variance:
  - 0.10% pa, [0.02%,0.18%], p=0.016
  - Weights: 24% '330', 46% Dream, 16% Adopt, 0% '0-0'
- Mantel-Haenszel:
  - 0.12% pa, [0.01%,0.23%], p=0.034
  - Weights: 3% '330', 36% Dream, 35% Adopt, 5% '0-0'

Random-effects

- The Q statistic is less than its expected value, and $I^2 = 0$
  - Hence the random-effects approach gives the same result

Odds ratio

There are several possible approaches using OR:

- (Inverse variance of log OR excludes 30 trials)
- Logistic reg. 1.427 [1.030, 1.977] p=0.0327
- Scoring 1.429 [1.031, 1.980] p=0.0320
- Conditional 1.426 [1.029, 1.975] p=0.0328

  (Exact using mid-P) p=0.0339
- Peto 1.428 [1.031, 1.979] p=0.0321
- (Mantel-Haenszel excludes 30 trials)

Bayesian approach

- Generalized linear mixed model
  - random Study effect
- Use “non-informative” priors
  - N(0; 10,000) for mean (on log-odds scale)
  - N(0; 10,000) for all Study effects (ditto)
- Fixed-effect for treatment
  - 0-0 trials don’t contribute
  - 1.45 with 95% credible interval [1.03,1.98]

Method is unimportant (as long as it is appropriate)

The choice of data drives the results

5. Trials with no events

- Agreement is reassuring, but 10 trials do not contribute
- Intuitively, they say something about relative risk and odds ratio:
  - An equal number of events (i.e. none) were observed
- But this intuition is based on prior expectation:
  - We expect that the risk is not actually zero, and that therefore there is an underlying relative risk
  - Without prior information, there is NO information about relative risk or odds ratio, because the observed risks are zero

Trials with no events (cont.)

There are three ways to address this problem

- Bayesian approach
  - Need priors, and different people (e.g. pharma companies and regulators) will naturally not agree
- "Continuity correction"
  - Adjust the zeroes slightly, on the grounds that they are expected to be non-zero
  - Introduces deliberate bias
- Combine trials to reduce sparsity
  - Danger of confounding and Simpson's Paradox
  - Standard approach for sparse contingency tables

Bayesian approach

- Random-effects model:
  - 0-0 trials contribute to the results
  - “non-informative” U(0; 10) for SD (τ)
  - Other priors as for Bayesian fixed-effects model
  - 1.52 [0.97, 2.36]
- So 0-0 trials increase the estimate a little, but the credible interval also widens
  - Surprising that they increase the estimate
  - I don’t know why

Adjusting the zeroes

- Add a small number to trials with zero events, to trials with zero in at least one arm, or to all the trials regardless
- Some software does this automatically, using the value 0.5
- Research indicated that this produces less biased results than ignoring the zero results (Refs 2 & 6)
  - However, they looked at risks no smaller than 10%
  - Here it is <1%; the 0.5s tends to swamp any real effects (Ref 7)
- E.g., adjust both counts in trials where one or both is zero
  - 0.5 reduces the combined estimate to 1.29 [0.95, 1.74] p=0.10
  - 0.1 to reduce swamping gives 1.40 [1.02, 1.93] p=0.040
  - 0.01 returns us almost to the result omitting the zero trials
- "Treatment-arm" correction, using 1/*N* from other treatment arm (Ref 8): 1.43 [1.03, 1.98] p=0.033

Pooling discriminately

- Combine trials to avoid having no events
  - Match trials by duration, treatment and rand. ratio

Trials with no MI events				Matched trials with some MI events			
Trial	Dur.	Treatments	Ratio	Trial	Dur.	Treatments	Ratio
095	26	Rosi+Ins vs Ins	2:1	082	26	Rosi+Ins vs Ins	2:1
234	26	Rosi+SU vs SU	2:1	079	26	Rosi±SU vs SU	2:1
331	52	Rosi vs Plac	2:1	330	52	Rosi vs Plac	3:1
009	24	Rosi+Met+Ins vs Ins	1:1	347	24	Rosi+Ins vs Ins	2:1
282	24	Rosi+Met vs SU+Met	1:1	284	24	Rosi+Met vs Met	1:1
369	26	Rosi vs SU	1:1	162	26	Rosi+SU vs SU	1:1
096	26	Rosi+SU vs SU	2:1	079	26	Rosi±SU vs SU	2:1
044	26	Rosi+Met vs Met	2:1	094	26	Rosi+Met vs Met	2:1
325	24	Rosi vs SU	1:1	143	24	Rosi+SU vs SU	1:1
004	24	Rosi±SU vs SU	3:1	132	24	Rosi+SU vs SU	4:1

Pooling (cont.)

- Effect of including the extra trials is minimal
  - estimates become 1.433 (1.035, 1.985) p=0.0303
- Surprising that this **increases** the estimate slightly
  - Non-intuitive (because of prior expectation)
  - Pooling reduces the underlying heterogeneity, and this increases the effect on the marginal scale
- Confirms that there is indeed no information about the odds ratio in these 10 trials (unless we adopt informative priors)

Result

- There is a signal from the MA (not evident in any study alone) that there are slightly more MIs for patients in the trials who received rosiglitazone
- Estimated odds ratio 1.43 (can also interpret as risk ratio)
- Could say that the rosiglitazone patients experienced a 43% higher risk of MI:
  - But it is confusing to talk of % of risk, which is itself often a %
  - Also, a risk difference would be more relevant to patients
- Do not analyse on risk scale to get risk difference:
  - Use the model to predict the risk difference
  - E,g. mildly diabetic patient with risk of 0.5% p.a. → 0.7% p.a. on average (risk difference of 0.2% or two in a thousand)
  - However, this average is over all comparators in this collection of trials, whereas patient knows own regimen

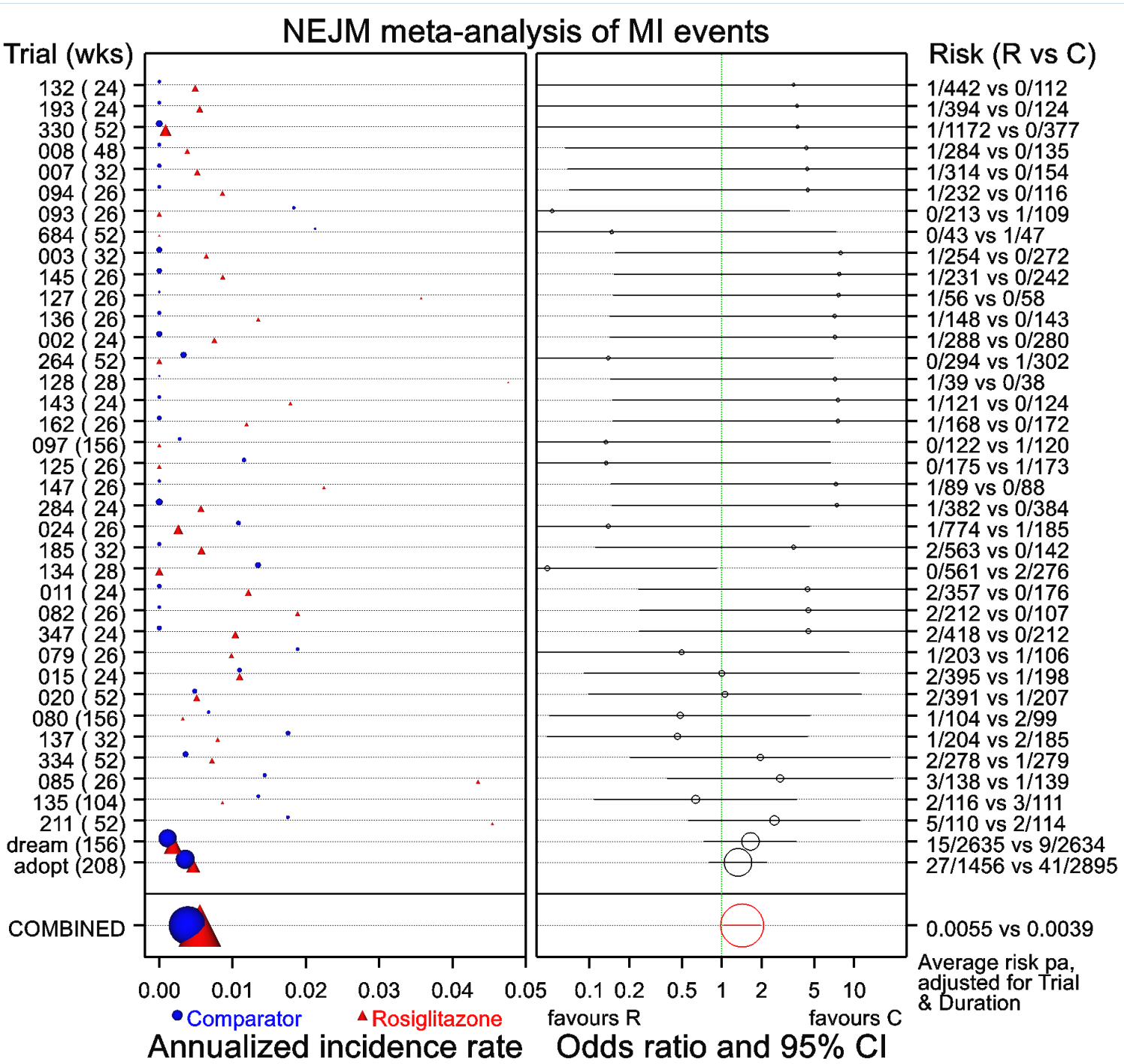
Prediction

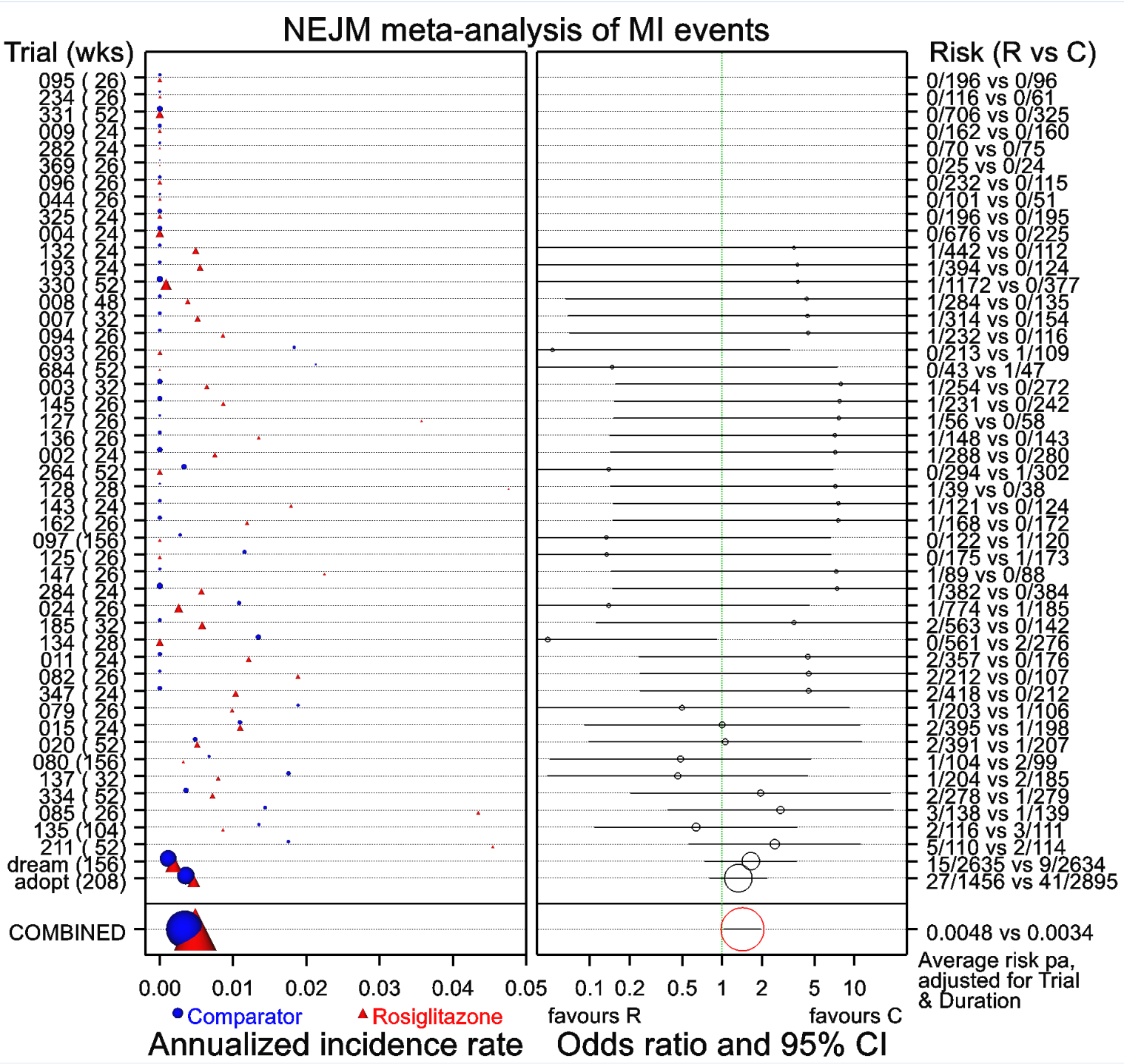
- Summary on the risk scale, averaging over the studies, adjusting for the differences
  - "Predicts" the average risk under adjusted conditions: equal numbers on each treatment
  - Stats calculated by LSMEANS /OM in the GLM procedure of SAS (not in GENMOD or LOGISTIC!)
- Fit logistic, adjusting for duration (offset), gives predicted average risk p.a. (from GenStat):
  - Comparator 0.34% (s.e. 0.050)
  - Rosiglitazone 0.48% (s.e. 0.059)

  Diff 0.14% (s.e. 0.066)
- It is important to average on the risk scale, not on the logistic scale (only option in SAS)

6. Graphical summary

- Graphical methods enhance interpretation
- Interval plot, or “Forest plot”, is standard for MA
  - Design to show weights of components
  - Add raw data to allow look-up
  - Give combined estimate(s) in margin
- Add panel of actual incidence, for context
- Add prediction as summary of combined effect


NEJM meta-analysis of MI events
Trial (wks)
Risk (R vs C)
132 (24)
193 (24)
330 (52)
008 (48)
007 (32)
094 (26)
093 (26)
684 (52)
003 (32)
145 (26)
127 (26)
136 (26)
002 (24)
264 (52)
128 (28)
143 (24)
162 (26)
097 (156)
125 (26)
147 (26)
284 (24)
024 (26)
185 (32)
134 (28)
011 (24)
082 (26)
347 (24)
079 (26)
015 (24)
020 (52)
080 (156)
137 (32)
334 (52)
085 (26)
135 (104)
211 (52)
dream (156)
adopt (208)
COMBINED
1/442 vs 0/112
1/394 vs 0/124
1/1172 vs 0/377
1/284 vs 0/135
1/314 vs 0/154
1/232 vs 0/116
0/213 vs 1/109
0/43 vs 1/47
1/254 vs 0/272
1/231 vs 0/242
1/56 vs 0/58
1/148 vs 0/143
1/288 vs 0/280
0/294 vs 1/302
1/39 vs 0/38
1/121 vs 0/124
1/168 vs 0/172
0/122 vs 1/120
0/175 vs 1/173
1/89 vs 0/88
1/382 vs 0/384
1/774 vs 1/185
2/563 vs 0/142
0/561 vs 2/276
2/357 vs 0/176
2/212 vs 0/107
2/418 vs 0/212
1/203 vs 1/106
2/395 vs 1/198
2/391 vs 1/207
1/104 vs 2/99
1/204 vs 2/185
2/278 vs 1/279
3/138 vs 1/139
2/116 vs 3/111
5/110 vs 2/114
15/2635 vs 9/2634
27/1456 vs 41/2895
0.0055 vs 0.0039
Average risk pa, adjusted for Trial & Duration
0.00 0.01 0.02 0.03 0.04 0.05
0.1 0.2 0.5 1 2 5 10
Comparator
Rosiglitazone
favours R
favours C
Annualized incidence rate
Odds ratio and 95% CI


NEJM meta-analysis of MI events
Trial (wks)
Risk (R vs C)
095 (26) 0/196 vs 0/96
234 (26) 0/116 vs 0/61
331 (52) 0/706 vs 0/325
009 (24) 0/162 vs 0/160
282 (24) 0/70 vs 0/75
369 (26) 0/25 vs 0/24
096 (26) 0/232 vs 0/115
044 (26) 0/101 vs 0/51
325 (24) 0/196 vs 0/195
004 (24) 0/676 vs 0/225
132 (24) 1/442 vs 0/112
193 (24) 1/394 vs 0/124
330 (52) 1/1172 vs 0/377
008 (48) 1/284 vs 0/135
007 (32) 1/314 vs 0/154
094 (26) 1/232 vs 0/116
093 (26) 0/213 vs 1/109
684 (52) 0/43 vs 1/47
003 (32) 1/254 vs 0/272
145 (26) 1/231 vs 0/242
127 (26) 1/56 vs 0/58
136 (26) 1/148 vs 0/143
002 (24) 1/288 vs 0/280
264 (52) 0/294 vs 1/302
128 (28) 1/39 vs 0/38
143 (24) 1/121 vs 0/124
162 (26) 1/168 vs 0/172
097 (156) 0/122 vs 1/120
125 (26) 0/175 vs 1/173
147 (26) 1/89 vs 0/88
284 (24) 1/382 vs 0/384
024 (26) 1/774 vs 1/185
185 (32) 2/563 vs 0/142
134 (28) 0/561 vs 2/276
011 (24) 2/357 vs 0/176
082 (26) 2/212 vs 0/107
347 (24) 2/418 vs 0/212
079 (26) 1/203 vs 1/106
015 (24) 2/395 vs 1/198
020 (52) 2/391 vs 1/207
080 (156) 1/104 vs 2/99
137 (32) 1/204 vs 2/185
334 (52) 2/278 vs 1/279
085 (26) 3/138 vs 1/139
135 (104) 2/116 vs 3/111
211 (52) 5/110 vs 2/114
dream (156) 15/2635 vs 9/2634
adopt (208) 27/1456 vs 41/2895
COMBINED 0.0048 vs 0.0034
Average risk pa, adjusted for Trial & Duration
0.00 0.01 0.02 0.03 0.04 0.05
0.1 0.2 0.5 1 2 5 10
Comparator
Rosiglitazone
favours R
favours C
Annualized incidence rate
Odds ratio and 95% CI

Context

- GSK patient-level meta-analysis (to FDA in 2006)
  - Separate analyses for different comparators
  - Many of the same trials, but not ADOPT or DREAM
  - Indication of some marginally raised incident rates
- GSK-commissioned observational study (2006)
  - 33,000 patients on oral anti-diabetics
  - Composite CV endpoint (MI and revascularization)
  - Hazard ratio 0.93 for Rosi vs other treatments
- Recent paper by Dahabreh in *Clinical Trials* (Ref 1)
  - Added new results from large trial (RECORD), and updated event counts
  - 1.43 estimate reduced to 1.33 – 1.23 (dep. on method)

References

1. Dahabreh IJ (2008). Meta-analysis of rare events: an update and sensitivity analysis of cardiovascular events in randomized trials of rosiglitazone. *Clinical Trials* **5**:116–120.
2. Gart JJ, Zweifel JR (1967). On the bias of various estimators of the logit and its variance with application to quantal assay. *Biometrika* **54**:181–187.
3. Greenland S, Salvan A (1990). Bias in the one-step method for pooling study results. *Statistics in Medicine* **9**:247–252.
4. Nissen SE, Wolski K (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine* 2007; 356.
5. Robbins J, Greenland S, Breslow N (1986). A general estimator for the variance of the Mantel-Haenszel odds ratio. *American Journal of Epidemiology* **124**:719–723.
6. Sankey SS, Weissfeld LA, Fine MJ, Kapoor W (1996). An assessment of the use of the continuity correction for sparse data in meta-analysis. *Communications in Statistics – Simulation and Computation* **25**:1031–1056.
7. Sutton AJ, Cooper NJ, Lambert PC, Jones DR, Abrams KR, Sweeting MJ (2002). Meta-analysis of rare and adverse event data. *Expert Review of Pharmacoeconomics and Outcomes Research* **2**:367–379.
8. Sweeting MJ, Sutton AJ, Lambert PC (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* **23**:1351–1375.