Exact meta-analysis for 2×2 tables with rare outcomes

Ken Rice & Spencer Hansen

JSM, August 2024

What's your talk about?

- 2×2 tables, from populations in which relevant odds ratios are heterogeneous
- Avandia! Recall Regina Liu on Tuesday, & see Yi Huang's talk later
- Exact intervals, for inference on the population(s) from whom we have data, a.k.a. *fixed effects* (FE) meta-analysis

Slides, papers and course materials at

tinyurl.com/fixef

Exact FE methods for 2×2 tables: motivation

- In 2007 Nissen & Wolski meta-analyzed data on Avandia (rosiglitazone maleate), looking at myocardial infarction (MI) and cardiovasular-related death (CVD) outcomes
- N&W found a significant effect of Avandia on MI...
- ...leading to many legal cases – see right. Avandia got a **black box warning**
- The N&W analysis used large-sample approximations, with small cell counts. Was that okay? Can it be trusted?

Exact FE methods for 2×2 tables: notation

We denote

	Outcome e.g. CVD 1	0	Total
Group 1 e.g. Avandia	X_i	$M_i - X_i$	M_i
Group 0 e.g. placebo	$T_i - X_i$	$N_i - T_i + X_i$	N_i
Total	T_i	$M_i + N_i - T_i$	$M_i + N_i$

,

for studies $i = 1, ..., k$.

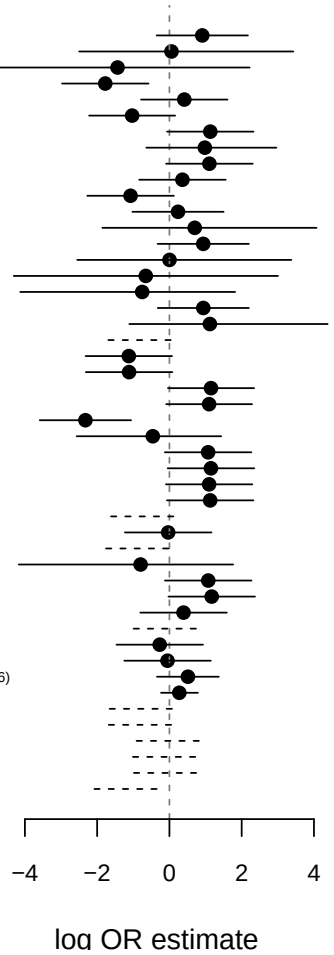
- Each group (i.e. M_i, N_i) at least moderate-size, but the bad outcome (side effect) is rare in any group; **expect small and sometimes zero counts**
- No confounding, and trial groups split roughly evenly
- Denote cell probabilities and odds ratio in population i as

$$\theta_{ij} = \mathbb{P}[\,\text{Outcome} = j | \text{Group} = i\,], \qquad \psi_i = \frac{\theta_{i1}}{1-\theta_{i1}} \frac{1-\theta_{i0}}{\theta_{i0}}$$

Exact FE methods for 2×2 tables: small counts!

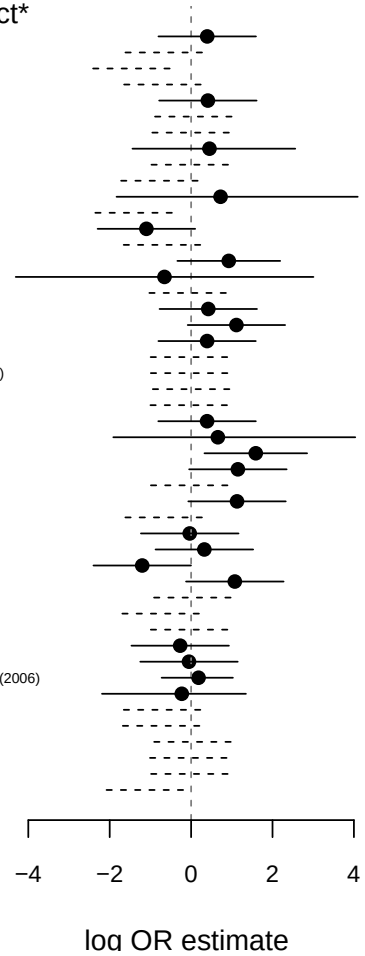
MI data

Study #	Paper/Abstract*	RSG M_k	RSG X_k	Ctrl N_k	Ctrl $T_k - X_k$
49653/011	Lebovitz et al (2001)	357	2	176	0
49653/020	Charbonnel et al (1999)*	391	2	207	1
49653/024	Phillips et al (2001)	774	1	185	1
49653/093	Jones et al (2003)	213	0	109	1
49653/094	Fonseca et al (2000)	232	1	116	0
100684		43	0	47	1
49653/143	Campbell et al (2004)*	121	1	124	0
49653/211		110	5	114	2
49653/284	Weissman et al (2005)	382	1	384	0
712753/008		284	1	135	0
AVM100264		294	0	302	1
BRL 49653C/185		563	2	142	0
BRL 49653/334		278	2	279	1
BRL 49653/347	Hollander et al (2005)*	418	2	212	0
49653/015	Wolffenbuttel et al (2000)	395	2	198	1
49653/079		203	1	106	1
49653/080	St John Sutton et al (2002)	104	1	99	2
49653/082	Raski et al (2001)	212	2	107	0
49653/085		138	3	139	1
49653/095		196	0	96	0
49653/097		122	0	120	1
49653/125	Vongthavaravat et al (2002)	175	0	173	1
49653/127	Garber et al (2001)*	56	1	58	0
49653/128		39	1	38	0
49653/134	Jones et al (2001)*	561	0	276	2
49653/135		116	2	111	3
49653/136		148	1	143	0
49653/145	Baksi et al (2004)	231	1	242	0
49653/147	Barnett et al (2003)	89	1	88	0
49653/162	Kerenyi et al (2004)	168	1	172	0
49653/234	Hamann et al (2003)*	116	0	61	0
49653/330		1172	1	377	0
49653/331	Ellis et al (2007)	706	0	325	0
49653/137		204	1	185	2
SB−712753/002	Bailey et al (2005)	288	1	280	0
SB−712753/003	Stewart et al (2006)	254	1	272	0
SB−712753/007		314	1	154	0
SB−712753/009	Home et al (2007)*	162	0	160	0
49653/132	Zhu et al (2003)	442	1	112	0
AVA100193		394	1	124	0
DREAM	The DREAM Investigators (2006)	2635	15	2634	9
ADOPT	Kahn et al (2007)	1456	27	2895	41
49653/044	Collegeville, PA (2001)*	101	0	51	0
49653/096		232	0	115	0
49653/282	Rood et al (2009)*	70	0	75	0
49653/369		25	0	24	0
49653/325	Rosenstock et al (2005)*	196	0	195	0
797620/004		676	0	225	0



CVD mortality data

Study #	Paper/Abstract*	RSG M_k	RSG X_k	Ctrl N_k	Ctrl $T_k - X_k$
49653/011	Lebovitz et al (2001)	357	1	176	0
49653/020	Charbonnel et al (1999)*	391	0	207	0
49653/024	Phillips et al (2001)	774	0	185	0
49653/093	Jones et al (2003)	213	0	109	0
49653/094	Fonseca et al (2000)	232	1	116	0
100684		43	0	47	0
49653/143	Campbell et al (2004)*	121	0	124	0
49653/211		110	3	114	2
49653/284	Weissman et al (2005)	382	0	384	0
712753/008		284	0	135	0
AVM100264		294	2	302	1
BRL 49653C/185		563	0	142	0
BRL 49653/334		278	0	279	1
BRL 49653/347	Hollander et al (2005)*	418	0	212	0
49653/015	Wolffenbuttel et al (2000)	395	2	198	0
49653/079		203	1	106	1
49653/080	St John Sutton et al (2002)	104	0	99	0
49653/082	Raski et al (2001)	212	1	107	0
49653/085		138	1	139	0
49653/095		196	1	96	0
49653/097		122	0	120	0
49653/125	Vongthavaravat et al (2002)	175	0	173	0
49653/127	Garber et al (2001)*	56	0	58	0
49653/128		39	0	38	0
49653/134	Jones et al (2001)*	561	1	276	0
49653/135		116	2	111	1
49653/136		148	2	143	0
49653/145	Baksi et al (2004)	231	1	242	0
49653/147	Barnett et al (2003)	89	0	88	0
49653/162	Kerenyi et al (2004)	168	1	172	0
49653/234	Hamann et al (2003)*	116	0	61	0
49653/330		1172	1	377	0
49653/331	Ellis et al (2007)	706	1	325	0
49653/137		204	0	185	1
SB−712753/002	Bailey et al (2005)	288	1	280	0
SB−712753/003	Stewart et al (2006)	254	0	272	0
SB−712753/007		314	0	154	0
SB−712753/009	Home et al (2007)*	162	0	160	0
49653/132	Zhu et al (2003)	442	1	112	0
AVA100193		394	1	124	0
DREAM	The DREAM Investigators (2006)	2635	12	2634	10
ADOPT	Kahn et al (2007)	1456	2	2895	5
49653/044	Collegeville, PA (2001)*	101	0	51	0
49653/096		232	0	115	0
49653/282	Rood et al (2009)*	70	0	75	0
49653/369		25	0	24	0
49653/325	Rosenstock et al (2005)*	196	0	195	0
797620/004		676	0	225	0



Exact FE methods for 2×2 tables: what's wrong?

N&W caused **trouble!**...

...and nerdy excitement, in some UW offices!

As well as MDs/industry/politicians pushing their own agendas:

- Shuster *et al* (2007, SiM) assert that random effects are "needed" – they use large-n methods, but exact version given by Gronsbell *et al* (2019, SiM)
- Assuming all $\psi_i = \psi$ (*common effect* model) can use stratified Fisher's exact test, but Tian *et al* (2009, Biostatistics) present exact CIs (also RR,RD)

✓✓✓
Handling small counts!

✗✗✗
The ψ_i aren't randomly sampled, nor are they homogeneous.

What to do instead?

- Inference **for the study populations at hand** (just like we use for trials)
- ... on some overall "average" odds ratio

For much more on these, see Rice, Higgins & Lumley (JRSSA, 2017) – on why/how fixed effect**S** (*plural!*) methods achieve this.

Our approach adapts this for rare events 2×2 tables, and exact inference.

Note: 'exact' means 'not anti-conservative' e.g. the guarantee that T1ER $\geq \alpha$ in Fisher's 'Exact' Test

What will you estimate/test?

Under homogeneity, it's standard to start with the hypergeometric distribution for each X_i, conditional on the row & column totals:

$$\mathbb{P}[\, X_i = x | M_i = m, N_i = n, T_i = t \,] = \frac{\binom{m}{x}\binom{n}{t-x}\psi_i^x}{\sum_{\max(0,t-n)\leq x' \leq \min(t,m)} \binom{m}{x'}\binom{n}{t-x'}\psi_i^{x'}}$$

where ψ_i is the odds ratio for study k.

Very helpfully, this is **exactly** equal to the convolution of a series of non-identical Bernoulli variables (Kou & Ying, 1996), with success probabilities

$$p_{ij}(\boldsymbol{\psi}) = \frac{1}{1 - \lambda_{ij}/\psi_i},$$

where $\boldsymbol{\psi}$ is the vector of subtable-specific odds ratios and λ_{ij} are roots of subtable-specific hypergeometric polynomials.

What will you estimate/test?

The well-known *conditional maximum likelihood estimator* is the solution to

$$\begin{aligned} X_i &= \sum_j p_{ij}(\widehat{\psi}_i), && \text{single table} \\ \sum_i X_i &= \sum_{i,j} p_{ij}(\widehat{\psi}_{CMLE}), && \text{multiple tables} \end{aligned}$$

Under heterogeneity, we **define overall odds ratio** ψ_F as the solution of

$$\sum_{i,j} p_{ij}(\psi_F \mathbf{1}_k) = \sum_{i,j} p_{ij}(\boldsymbol{\psi})$$

... and $\widehat{\psi}_{CMLE}$ estimates ψ_F consistently, as number of subtables k increases.

Big idea: use $\sum_i X_i$ for inference about mean of the p_{ij}, and transform this to get inference on ψ_F.

What is the distribution of $\sum_i X_i$?

The point null $\psi_F = \psi_0$ doesn't specify a distribution for $\sum_i X_i$ – it only specifies the mean. But...

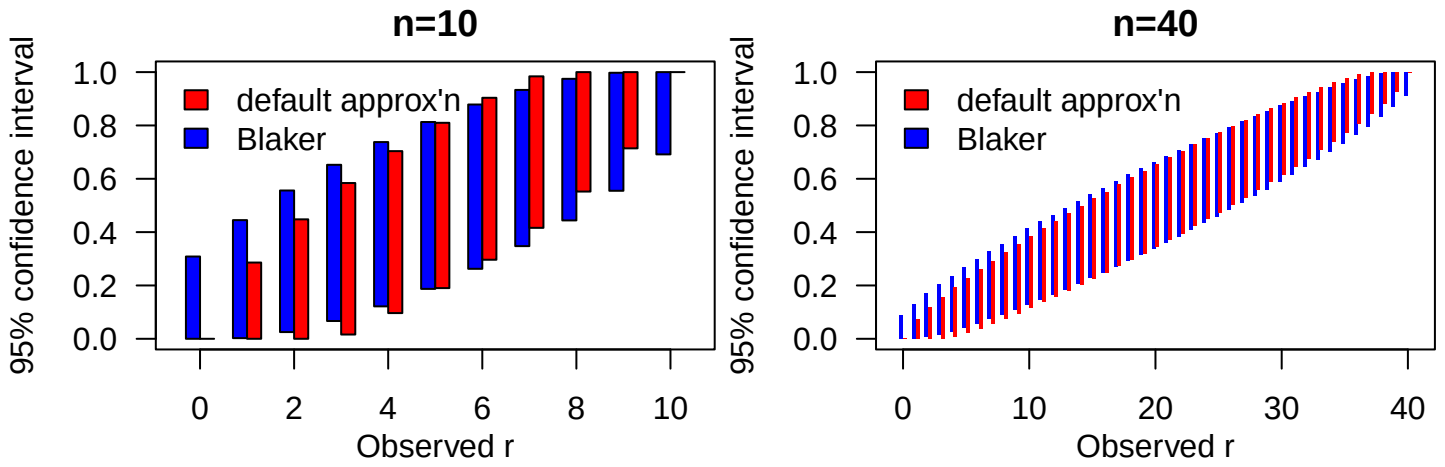
Among **all** convolutions of n Bernoullis with a specific overall mean $n\theta$, the 'pure' Binomial(n, θ) has the lightest tails.

Hoeffding (Annals, 1956)

So exact inference using $\sum_i X_i$ for its mean under the Binomial under homogeneous p_{ij} **also** provides exact inference under heterogeneous p_{ij}, we which we get with heterogeneous ψ_i.

Exact inference for binomial mixtures

Based on earlier work, the Blaker approaches (for r successes out of n) work well for exact tests **and** confidence intervals



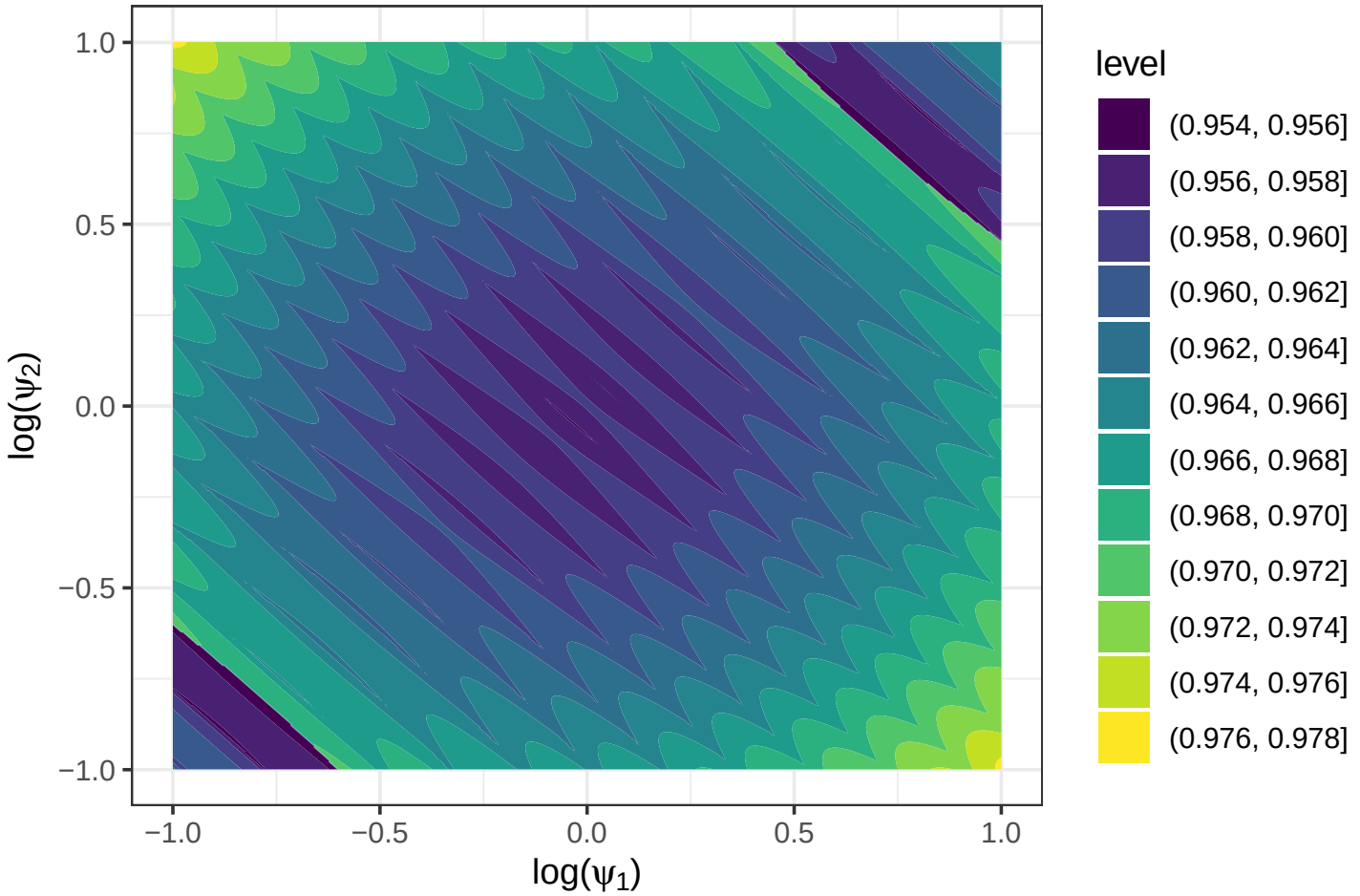
- We use a Blaker interval around $\sum_i X_i / \sum_i M_i + N_i$, which is exact when all p_{ij} are identical
- By Hoeffding, it is conservative as a statement of $\sum X_i$'s extremity under any ψ **that have the same overall** ψ_F
- Hence, after transformation to the ψ_F scale, gives exact inference for ψ_F

Does it work? (Yes! Of course!)

- Here we consider $k = 2$ 2×2 tables, each with $M_i{=}N_i{=}(500{,}500)$ and $T_i = (25, 25)$
- For each $\{\psi_1, \psi_2\}$, coverage rate is for ψ_F, the solution to

$$\sum_{i,j} p_{ij}(\psi_F) = \sum_{i,j} p_{ij}(\psi_i)$$

- Calculations use complete enumeration



‘Exact’ means a bit conservative: but how much?

Can learn about this comparing the variance of $\sum X_i$ (i.e. $\sum_{ij} p_{ij}(\psi_i)(1 - p_{ij}(\psi_i))$ to the pure binomial variance.

The critical quantity is

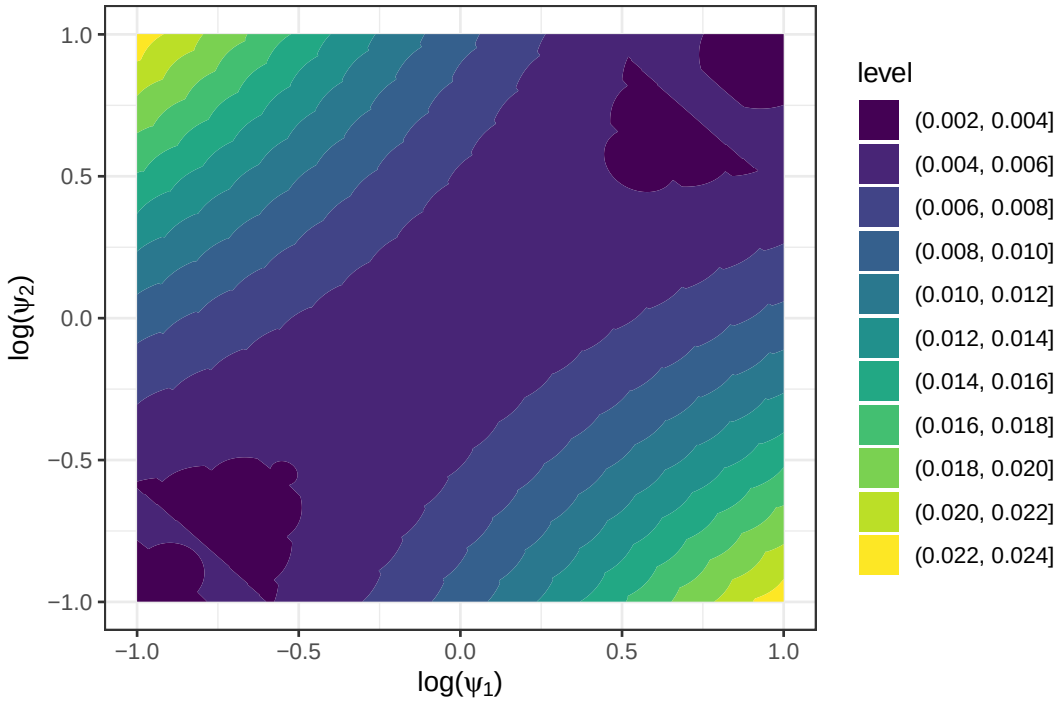
$$\delta^2 = \sum_{i,j} \frac{1}{\sum_i U_i} \left(p_{ji}(\psi_i) - \frac{\sum_{k,j} p_{ji}(\psi_F)}{\sum_i U_i} \right)^2, \text{ where } U_i = min(T_i, M_i).$$

- δ indicates **approximately** how much bigger the intervals are than they need to be – conservatism due to using pure binomial case **and** due to ψ_i being heterogeneous
- Doesn’t describe conservatism (under pure binomial) because we want exact intervals from a discrete sample space – but can simply compute that, with complete enumeration.

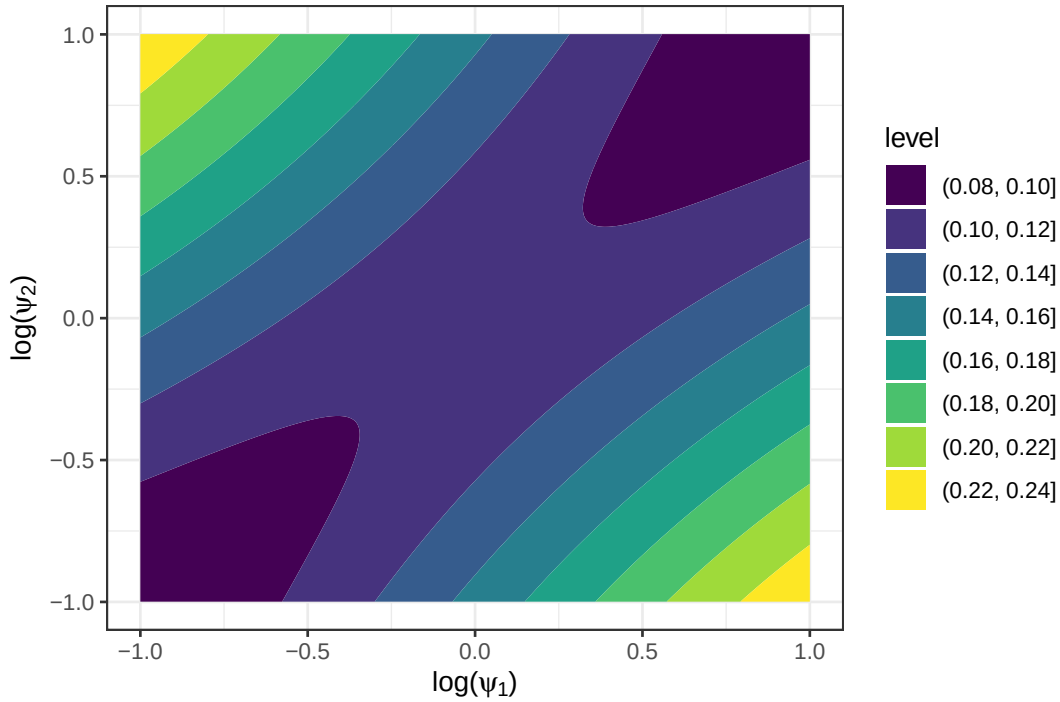
‘Exact’ means a bit conservative: but how much?

Again for k=2, M_i=N_i=(500,500), T_i=(25,25), showing how δ tracks excess coverage:

Excess coverage



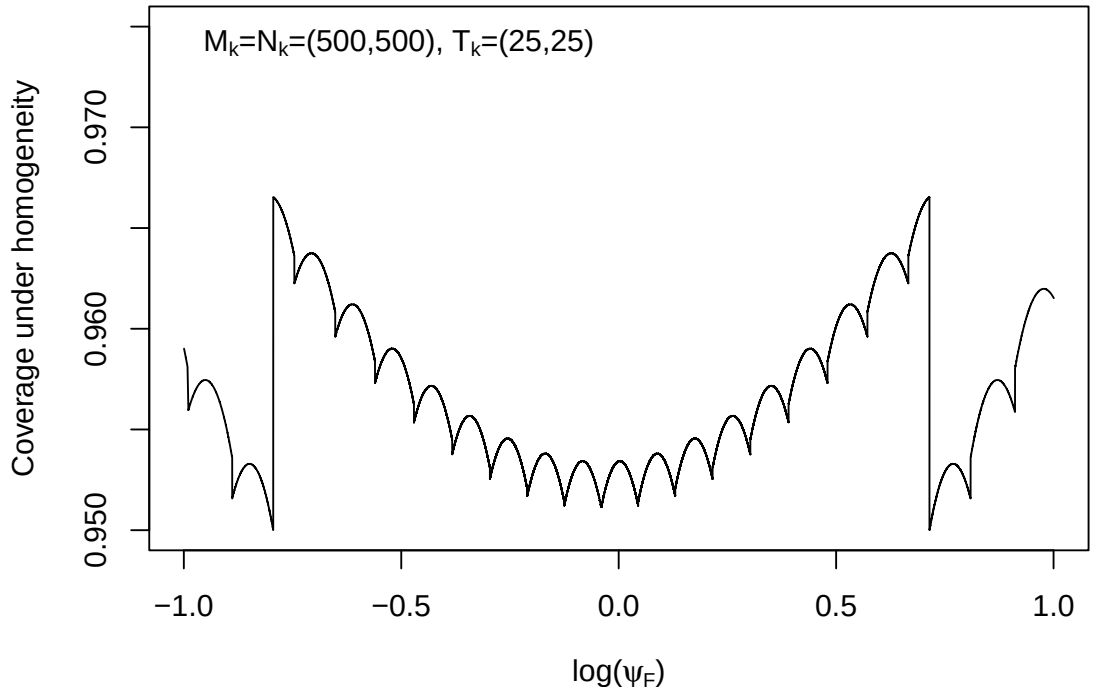
δ



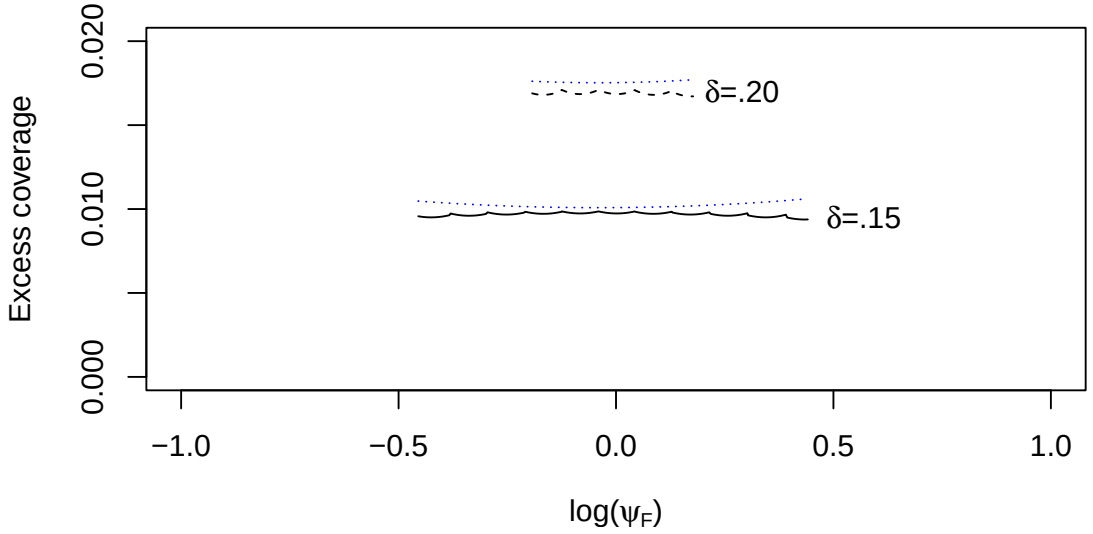
‘Exact’ means a bit conservative: but how much?

Turning δ’s excess width into approximate excess coverage, beyond what we get from pure binomial:

Coverage under pure binomial

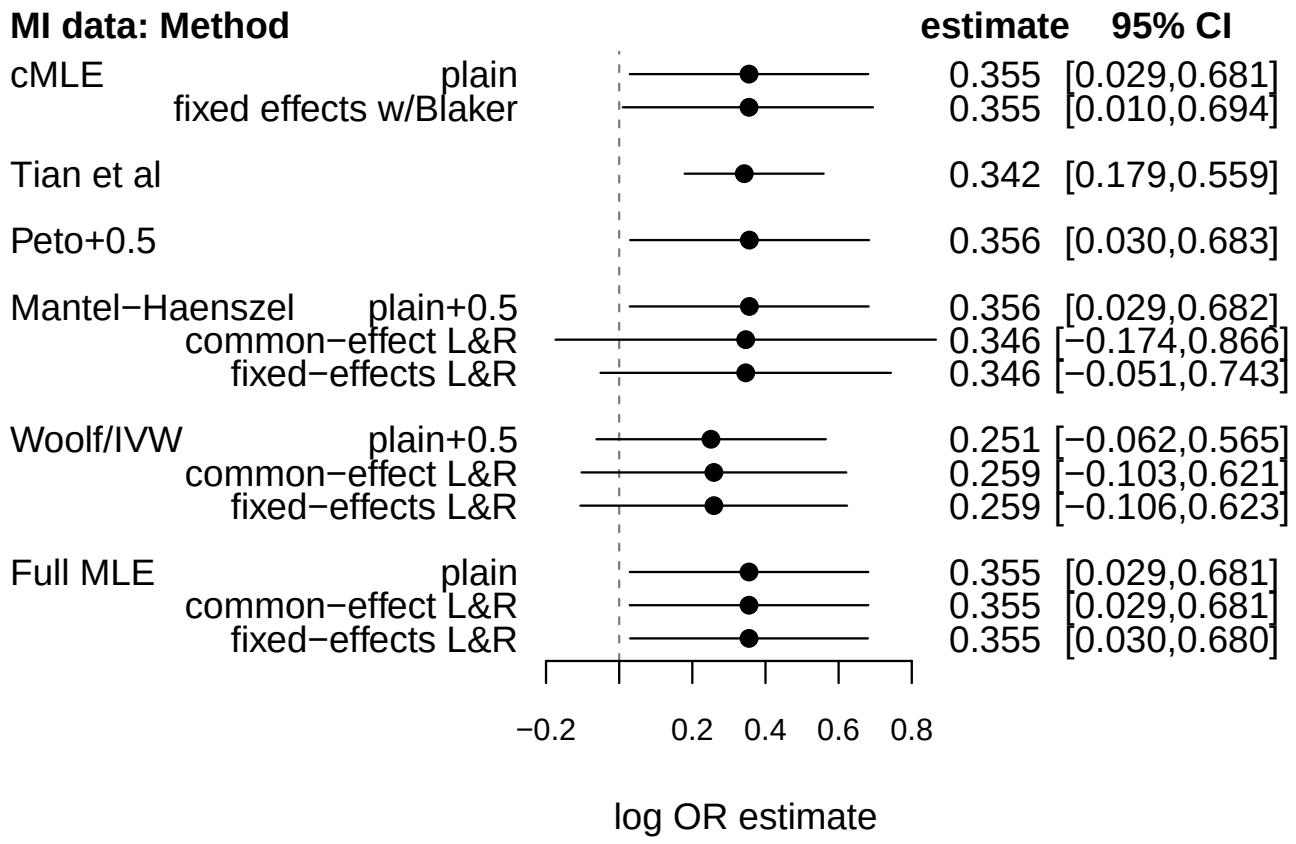


Actual/approximate excess coverage



Weren't you talking about Avandia?

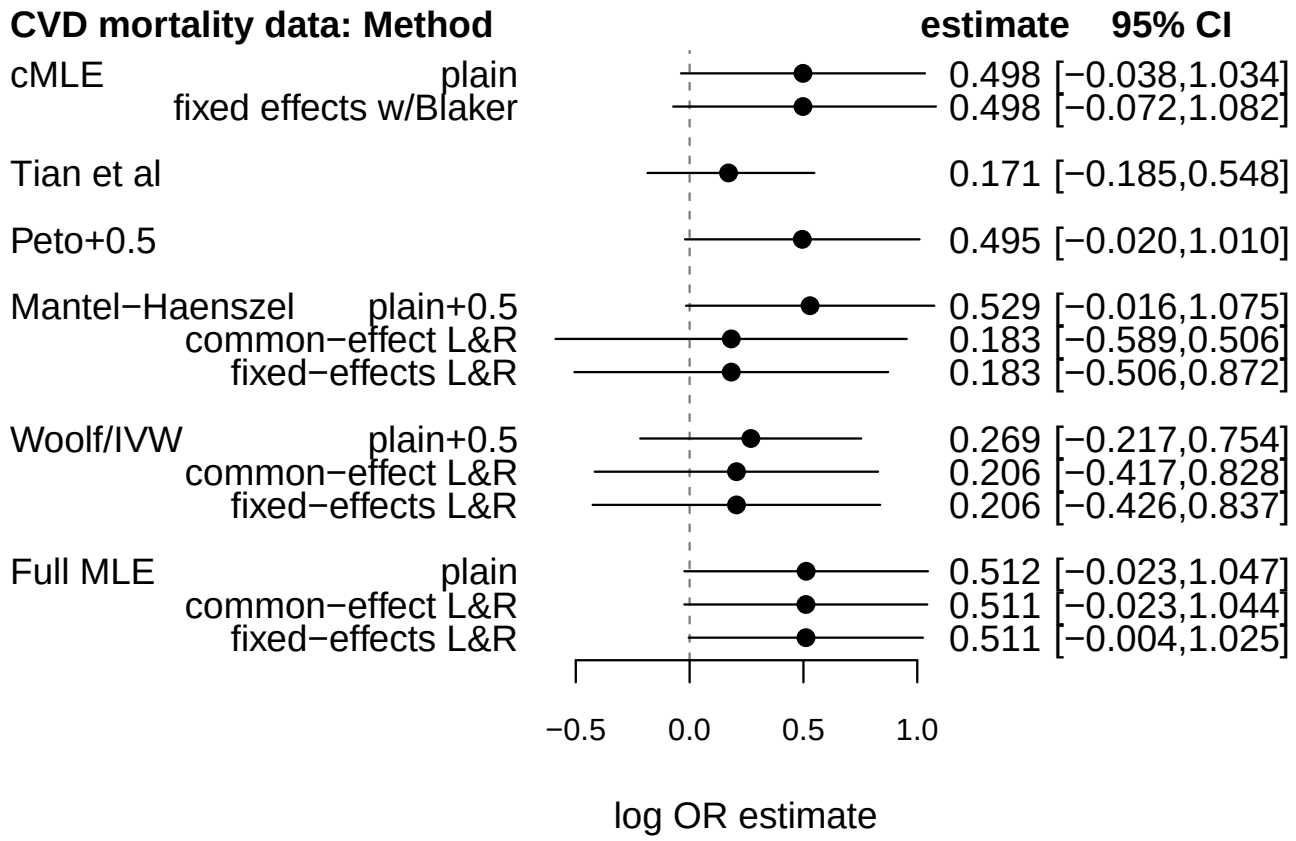
Comparing the exact method (cMLE fixed effects w/Blaker) to others:



$\widehat{\delta} = 0.18$, 0.14 under homogeneous ψ_i, suggests 96.4% (95.9%) coverage

Weren't you talking about Avandia?

Same comparisons for the other endpoint:



$\widehat{\delta} = 0.18$, 0.10 under homogeneous ψ_i, suggests 96.5% (95.6%) coverage

Are you going to stop now?

In summary:

- Nissen & Wolski analysis was fine - for FE inference on ORs
- Hansen & Rice (SiM 2023) gives exact methods for meta-analysis of 2×2 contingency tables, for inference on the conditional MLE odds ratio ψ_F
- The conservatism of the exact inference can be usefully approximated via δ^2
- Method is appropriate for any application where events are rare and heterogeneity is plausible. (For *common* events, can be very conservative)
- Method is specific to ORs, for which double-zero tables don't contribute

Thank you! (Particularly Yong, and Spencer)

tinyurl.com/fixef