Inferring replicability from Cohrane reviews

Daniel Yekutieli

Statistics and OR
Tel Aviv University

2 December 2014

Plan

1. Cochrane collaboration reviews
2. Testing framework for providing confidence statements regarding distribution of effect in new study
3. Implementation of mean most powerful tests
4. Results

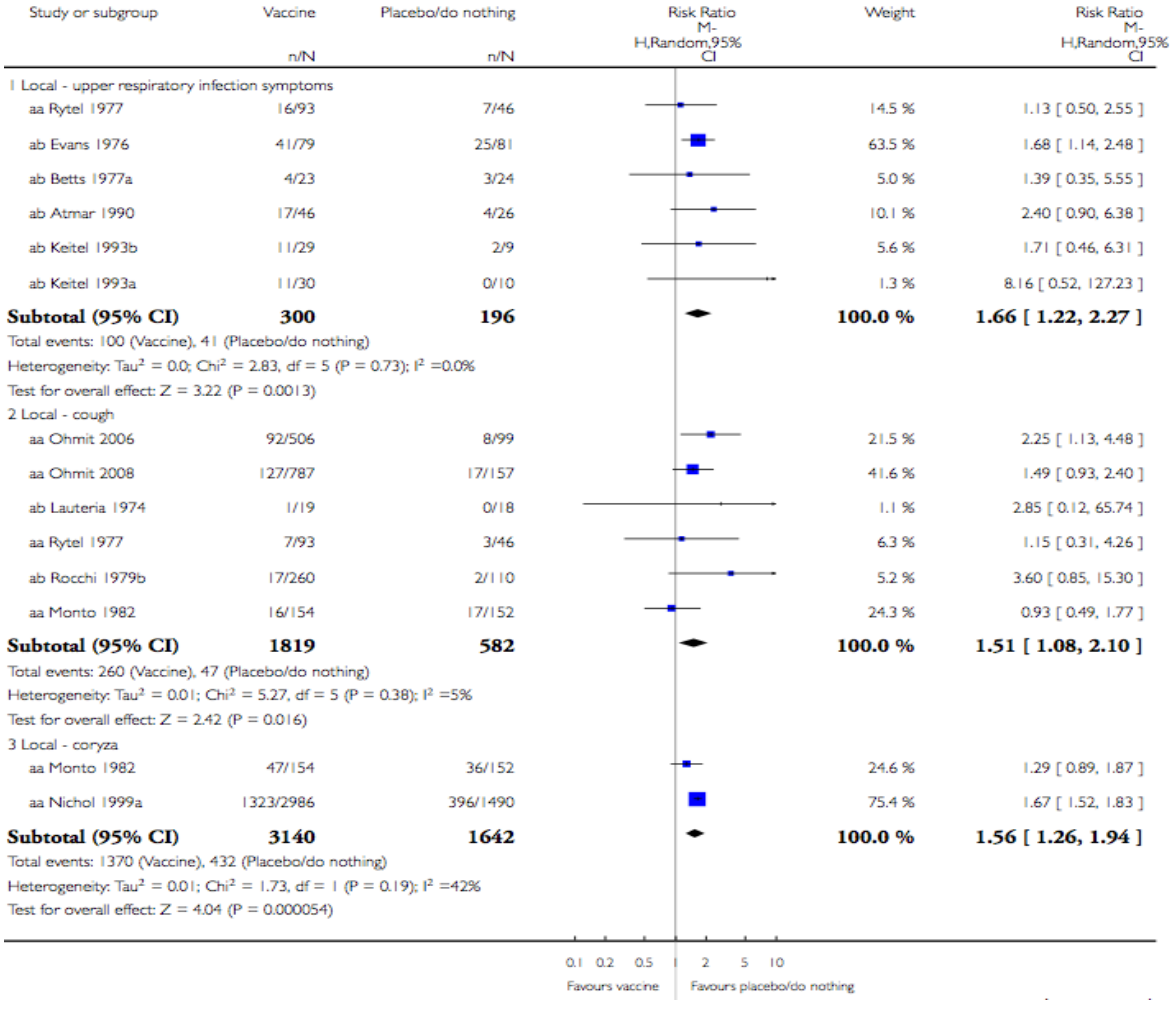
Influenza reviews – Outcome 4: local harms

Analysis 2.4. Comparison 2 Live aerosol vaccine versus placebo or 'do nothing', Outcome 4 Local harms.

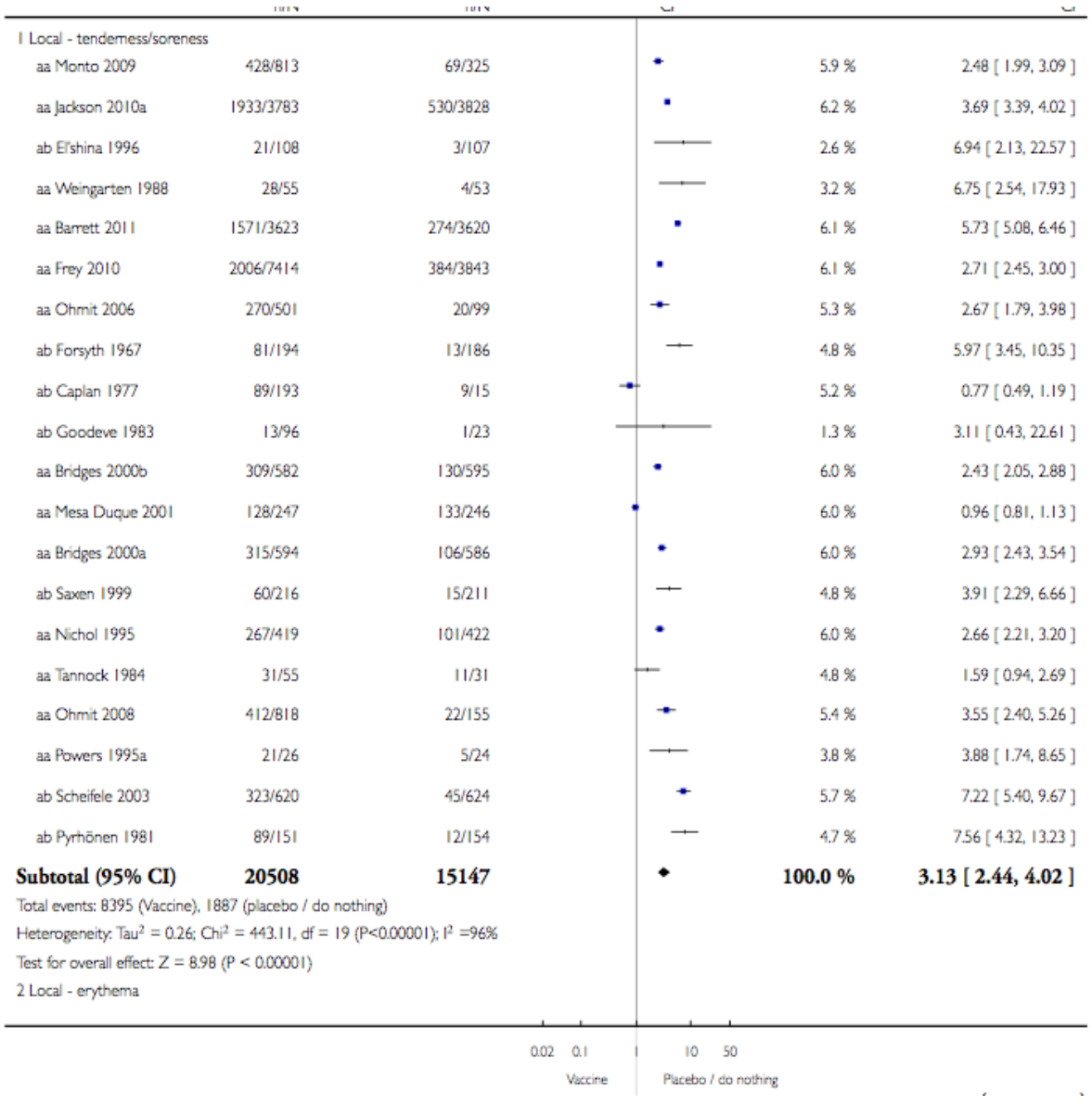
Review: Vaccines for preventing influenza in healthy adults

Comparison: 2 Live aerosol vaccine versus placebo or 'do nothing'

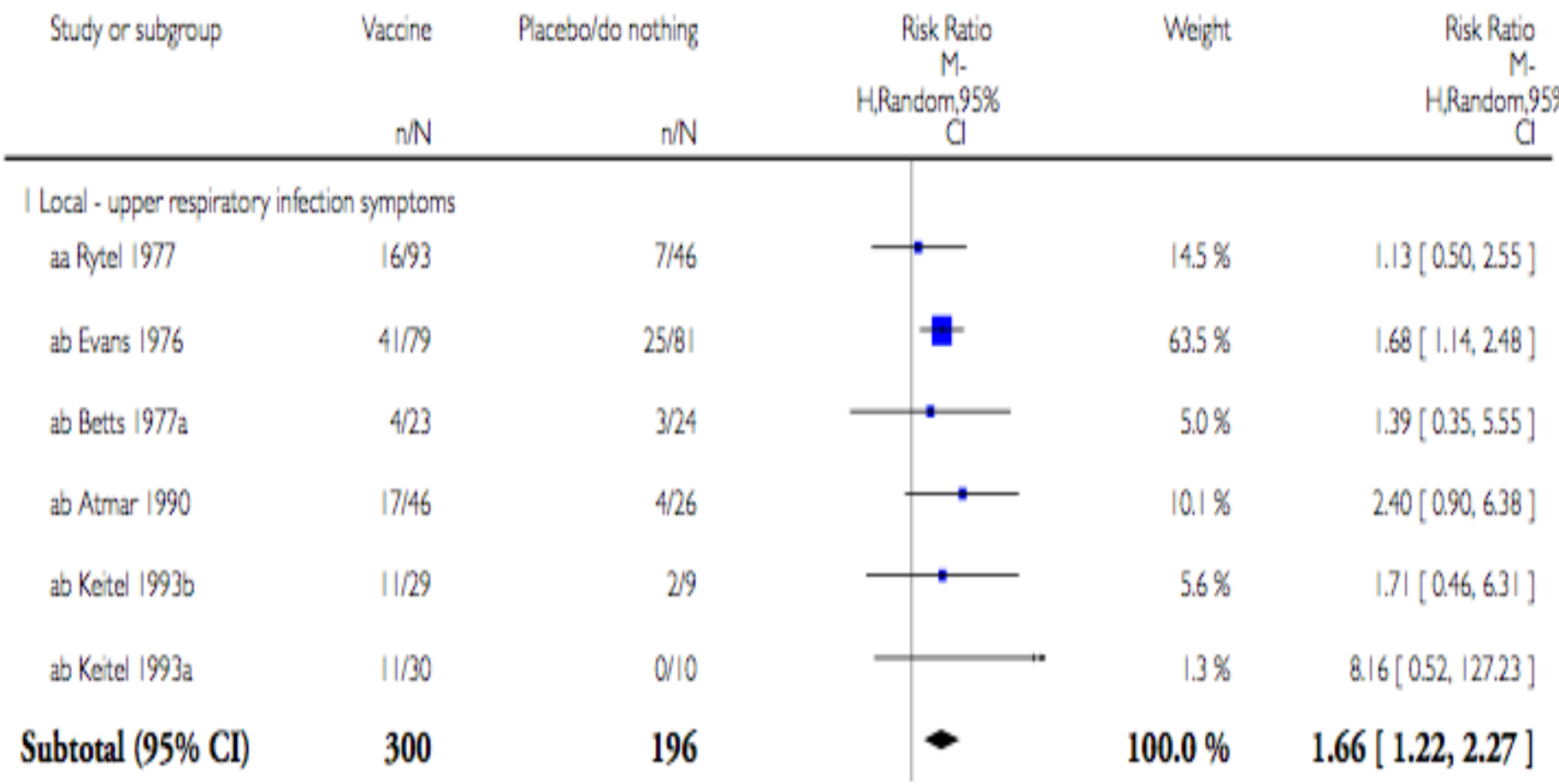
Outcome: 4 Local harms



Tenderness

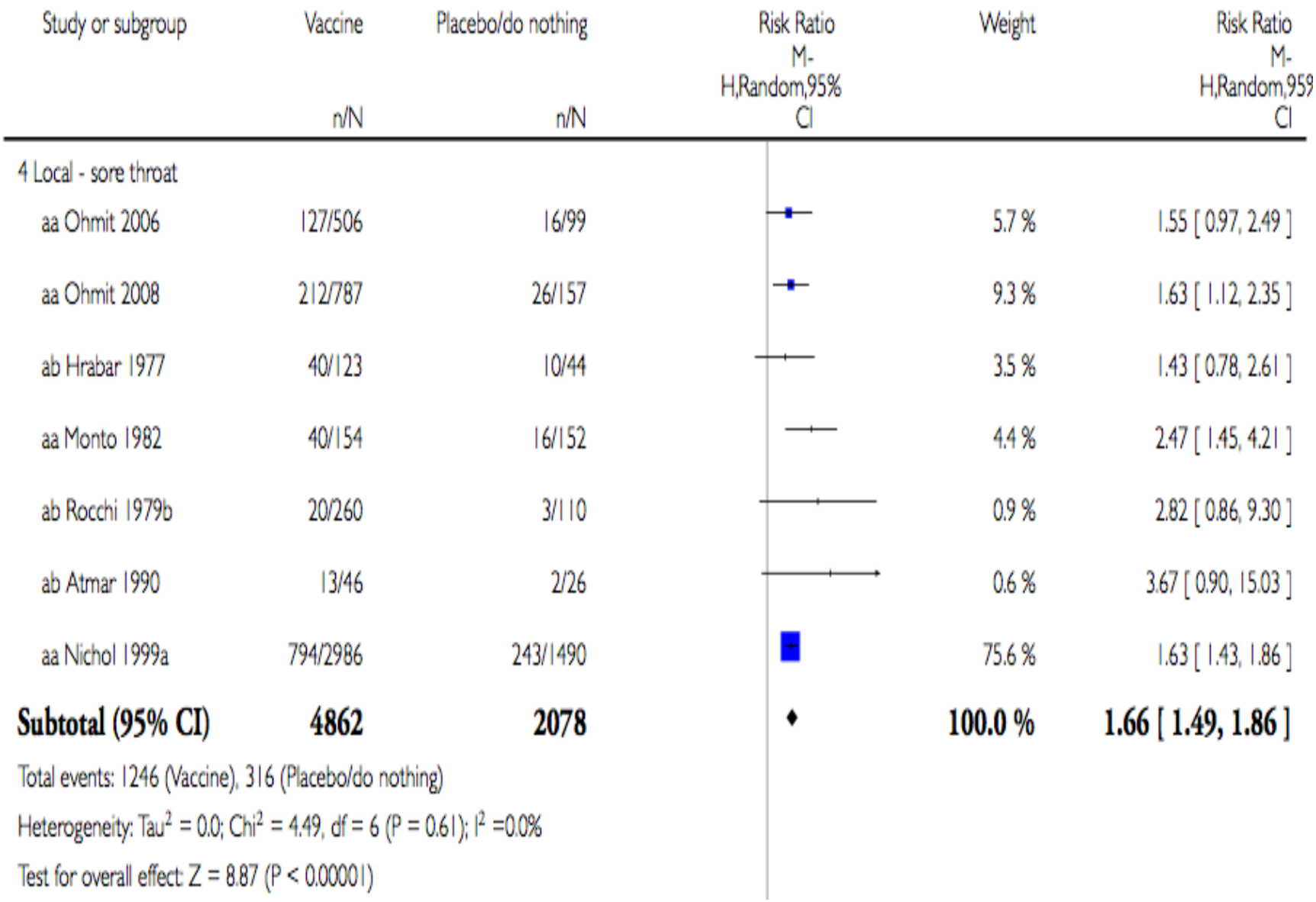

1 Local - tenderness/soreness
aa Monto 2009 428/813 69/325 5.9 % 2.48 [1.99, 3.09]
aa Jackson 2010a 1933/3783 530/3828 6.2 % 3.69 [3.39, 4.02]
ab El'shina 1996 21/108 3/107 2.6 % 6.94 [2.13, 22.57]
aa Weingarten 1988 28/55 4/53 3.2 % 6.75 [2.54, 17.93]
aa Barrett 2011 1571/3623 274/3620 6.1 % 5.73 [5.08, 6.46]
aa Frey 2010 2006/7414 384/3843 6.1 % 2.71 [2.45, 3.00]
aa Ohmit 2006 270/501 20/99 5.3 % 2.67 [1.79, 3.98]
ab Forsyth 1967 81/194 13/186 4.8 % 5.97 [3.45, 10.35]
ab Caplan 1977 89/193 9/15 5.2 % 0.77 [0.49, 1.19]
ab Goodeve 1983 13/96 1/23 1.3 % 3.11 [0.43, 22.61]
aa Bridges 2000b 309/582 130/595 6.0 % 2.43 [2.05, 2.88]
aa Mesa Duque 2001 128/247 133/246 6.0 % 0.96 [0.81, 1.13]
aa Bridges 2000a 315/594 106/586 6.0 % 2.93 [2.43, 3.54]
ab Saxen 1999 60/216 15/211 4.8 % 3.91 [2.29, 6.66]
aa Nichol 1995 267/419 101/422 6.0 % 2.66 [2.21, 3.20]
aa Tannock 1984 31/55 11/31 4.8 % 1.59 [0.94, 2.69]
aa Ohmit 2008 412/818 22/155 5.4 % 3.55 [2.40, 5.26]
aa Powers 1995a 21/26 5/24 3.8 % 3.88 [1.74, 8.65]
ab Scheifele 2003 323/620 45/624 5.7 % 7.22 [5.40, 9.67]
ab Pyrhönen 1981 89/151 12/154 4.7 % 7.56 [4.32, 13.23]
Subtotal (95% CI) 20508 15147 100.0 % 3.13 [2.44, 4.02]
Total events: 8395 (Vaccine), 1887 (placebo / do nothing)
Heterogeneity: Tau² = 0.26; Chi² = 443.11, df = 19 (P<0.00001); I² =96%
Test for overall effect: Z = 8.98 (P < 0.00001)
2 Local - erythema
0.02 0.1 1 10 50
Vaccine
Placebo / do nothing

Upper respiratory infection symptoms



Study or subgroup	Vaccine n/N	Placebo/do nothing n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Local - upper respiratory infection symptoms					
aa Rytel 1977	16/93	7/46		14.5 %	1.13 [0.50, 2.55]
ab Evans 1976	41/79	25/81		63.5 %	1.68 [1.14, 2.48]
ab Betts 1977a	4/23	3/24		5.0 %	1.39 [0.35, 5.55]
ab Atmar 1990	17/46	4/26		10.1 %	2.40 [0.90, 6.38]
ab Keitel 1993b	11/29	2/9		5.6 %	1.71 [0.46, 6.31]
ab Keitel 1993a	11/30	0/10		1.3 %	8.16 [0.52, 127.23]
Subtotal (95% CI)	**300**	**196**		**100.0 %**	**1.66 [1.22, 2.27]**

Sore throat



Study or subgroup	Vaccine n/N	Placebo/do nothing n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
4 Local - sore throat					
aa Ohmit 2006	127/506	16/99		5.7 %	1.55 [0.97, 2.49]
aa Ohmit 2008	212/787	26/157		9.3 %	1.63 [1.12, 2.35]
ab Hrabar 1977	40/123	10/44		3.5 %	1.43 [0.78, 2.61]
aa Monto 1982	40/154	16/152		4.4 %	2.47 [1.45, 4.21]
ab Rocchi 1979b	20/260	3/110		0.9 %	2.82 [0.86, 9.30]
ab Atmar 1990	13/46	2/26		0.6 %	3.67 [0.90, 15.03]
aa Nichol 1999a	794/2986	243/1490		75.6 %	1.63 [1.43, 1.86]
Subtotal (95% CI)	**4862**	**2078**		**100.0 %**	**1.66 [1.49, 1.86]**

Total events: 1246 (Vaccine), 316 (Placebo/do nothing)

Heterogeneity: $Tau^2 = 0.0$; $Chi^2 = 4.49$, df = 6 (P = 0.61); $I^2 = 0.0\%$

Test for overall effect: Z = 8.87 (P < 0.00001)

Setup and goal

- The data is the of effect estimators from K studies, $\hat{\theta}_1 \cdots \hat{\theta}_K$
- The study effects, $\theta_1 \cdots \theta_K$, are iid samples from an unknown effect population distribution $\pi(\theta)$
- Our goal is to provide inference regarding the distribution of θ_0, the treatment effect in a new treatment group assumed to be also independently sampled from $\pi(\theta)$

Inferences regarding distribution of effect of new treatment

1. Confidence interval for q_p, p'th quantile of $\pi(\theta)$:

$$a, b \in \mathbb{R} \text{ such that } \Pr_{\hat{\theta}_1 \cdots \hat{\theta}_K}(q_p \in [a, b]) \geq 1 - \alpha$$

2. Confidence interval for $p_A = Pr_{\theta_0 \sim \pi}(\theta_0 \in A)$:

$$0 \leq \hat{p}_a \leq \hat{p}_b \leq 1 \text{ such that } \Pr_{\hat{\theta}_1 \cdots \hat{\theta}_K}(p_A \in [\hat{p}_a, \hat{p}_b]) \geq 1 - \alpha$$

3. Lower bound for predictive probability that $\theta_0 \in A$

Deriving right tailed $1-\alpha$ confidence intervals for q_p

Algorithm:

1. For all $\tilde{a} \in \mathbb{R}$ we define

$$\Omega_0(\tilde{a}) = \{\pi : q_p(\pi) \leq \tilde{a}\} \text{ and } \Omega_1(\tilde{a}) = \{\pi : \tilde{a} < q_p(\pi)\}$$

2. and we use $\hat{\theta}_1 \cdots \hat{\theta}_K$ for level α test of

$$H_0(\tilde{a}) : \pi \in \Omega_0(\tilde{a}) \text{ vs. } H_1(\tilde{a}) : \pi \in \Omega_1(\tilde{a})$$

3. Set $a = \inf\{\tilde{a} : \text{ such that } H_0(\tilde{a}) \text{ is accepted}\}$

Denote the true π, $\bar{\pi}$ with p'th quantile $\bar{q}_p = q_p(\bar{\pi})$.
Therefore $H_0(\bar{q}_p)$ is true, thus it is accepted with probability $\geq 1-\alpha$, and if it is accepted then $a \leq \bar{q}_p$. Thus we get,

$$\Pr_{\hat{\theta}_1 \cdots \hat{\theta}_K}(a \leq \bar{q}_p) \geq \Pr_{\hat{\theta}_1 \cdots \hat{\theta}_K}(H_0(\bar{q}_p) \text{ is accepted }) \geq 1-\alpha.$$

Right tailed CI for median in the no-noise case with $K = 8$

Observed sequence of effects: $-2, -0.5, 1.1, 2.4, 3.2, 4.9, 7.3, 7.5$

- Test for $q_{1/2}(\pi) \leq \tilde{a}$: let $n(\tilde{a}) = \#\{\theta_k \leq \tilde{a}\}$, accept $H_0(\tilde{a})$ if

$$0.05 \ \leq \ \Pr\{X \leq n(\tilde{a})\} \ \text{ for } \ X \sim \mathit{Binomial}(8, 0.5).$$

- R function: $\mathit{pbinom}(c(8, 7, 6, 5, 4, 3, 2, 1, 0), 8, 0.5)$

$$1.000, 0.996, 0.965, 0.855, 0.637, 0.363, 0.145, 0.035, 0.004$$

- Thus we accept $H_0(\tilde{a})$ for $\tilde{a}$ with $3 \leq n(\tilde{a})$,

$$1.1 = \inf\{\tilde{a} : \ \text{such that } H_0(\tilde{a}) \text{ is accepted}\}$$

$\Rightarrow$ right tailed 0.95 CI for median of π is $[1.1, \infty]$

Deriving right tailed $1-\alpha$ confidence intervals for p_A

Algorithm:

1. For all $\tilde{p}_A \in [0,1]$ we define

$$\Omega_0(\tilde{p}_A) = \{\pi : p_A(\pi) \leq \tilde{p}_A\} \text{ and } \Omega_1(\tilde{p}_A) = \{\pi : \tilde{p}_A < p_A(\pi)\}$$

2. and we use $\hat{\theta}_1 \cdots \hat{\theta}_K$ for level α test of

$$H_0(\tilde{p}_A) : \pi \in \Omega_0(\tilde{p}_A) \text{ vs. } H_1(\tilde{p}_A) : \pi \in \Omega_1(\tilde{p}_A)$$

3. Set $\hat{p}_A = \inf\{\tilde{p}_A : \text{ such that } H_0(\tilde{p}_A) \text{ is accepted}\}$

For true $\bar{\pi}$ with $\bar{p}_A = \Pr_{\theta_0 \sim \bar{\pi}}(\theta_0 \in A)$, $H_0(\bar{p}_A)$ is true, it is accepted with probability $\geq 1-\alpha$ and if it is accepted then $\hat{p}_A \leq \bar{p}_A$. Thus we get

$$\Pr_{\hat{\theta}_1 \cdots \hat{\theta}_K}(\hat{p}_A \leq \bar{p}_A) \geq \Pr_{\hat{\theta}_1 \cdots \hat{\theta}_K}(H_0(\bar{p}_A) \text{ is accepted }) \geq 1-\alpha.$$

Right tailed CI for $p_{[0,\infty]}$ in the no-noise case with $K = 8$

Same sequence of effects: $-2, -0.5, 1.1, 2.4, 3.2, 4.9, 7.3, 7.5$

- Test for $p_{[0,\infty]}(\pi) \leq \tilde{p}_A$: let $n_{[0,\infty]} = \#\{k : 0 \leq \theta_k\}$, accept $H_0(\tilde{p}_A)$ if

$$0.05 \ \leq \ \Pr\{n_{[0,\infty]} \leq X\} \ \text{ for } \ X \sim Binomial(8, \tilde{p}_A)$$

- R function: $1 - pbinom(6, 8, c(0.8, 0.6, 0.530, 0.529, 0.50)), 4)$

$$0.5033, 0.1064, 0.0504, 0.0498, 0.0352$$

- We accept $H_0(\tilde{p}_A)$ with $0.530 \leq \tilde{p}_A$

$\Rightarrow$ right tailed 0.95 CI for $\Pr_{\theta_0 \sim \pi}(0 \leq \theta_0)$ is $[0.530, 1]$

Deriving lower bound for predictive probability that $0 \leq \theta_0$

- $\theta_0 \geq 0$ occurs, independently of $\hat{\theta}_1 \cdots \hat{\theta}_K$, with prob $p_0(\bar{\pi})$.
- To bound the prob that this event occurs we use the lower bound $\hat{p}_0 = 0.530$ that was based $\hat{\theta}_1 \cdots \hat{\theta}_K$.
- Let $Acc(\bar{\pi})$ denote the event that null hypothesis corresponding to $\bar{\pi}$, used for constructing the CI, is accepted
- $Acc(\bar{\pi})$ occurs with probability ≥ 0.95 and if it occurs then $0.530 \leq p_0(\bar{\pi})$

$$\begin{aligned}
\Pr_{\theta_0;\hat{\theta}_1\cdots\hat{\theta}_K}\{0 \leq \theta_0\} &\geq \Pr_{\theta_0;\hat{\theta}_1\cdots\hat{\theta}_K}\{\theta_0 \in A,\ Acc(\bar{\pi})\} \\
&= \Pr_{\theta_0}\{\theta_0 \in A |\ Acc(\bar{\pi})\} \cdot \Pr_{\hat{\theta}_1\cdots\hat{\theta}_K}\{Acc(\bar{\pi})\} \\
&\geq 0.530 \cdot 0.95 = 0.5035
\end{aligned}$$

Some comments and a question

(a) Left tailed CI's are derived similarly.

(b) $1-\alpha/2$ left tailed and right tailed CI's yield a two-tailed $1-\alpha$ CI.

(c) The tests used for CI's for quantiles and probabilities are the same:

1. For $A \subset \mathbb{R}$ and $\tilde{p} \in (0,1)$ we define two sets of distributions:

$$\Omega_0(\tilde{p}, A) = \{\pi : p_A(\pi) \leq \tilde{p}\} \text{ and } \Omega_1(\tilde{p}, A) = \{\pi : \tilde{p} < p_A(\pi)\}$$

2. We use the statistic $\#\{\theta_k \in A\}$ to test

$$H_0 : \pi \in \Omega_0(\tilde{p}, A) \text{ vs. } H_1 : \pi \in \Omega_1(\tilde{p}, A)$$

Q: How do we test these null hypotheses with $\hat{\theta}_1 \cdots \hat{\theta}_K$?

Mean most powerful tests

General framework for testing complex hypotheses presented in Yekutieli (2014). This is how it can be adapted to this application:

- The parameter is $\pi \in \Omega$ with prior distribution $\mathcal{D}(\pi)$
- the data is $\hat{\theta} = (\hat{\theta}_1 \cdots \hat{\theta}_K)$ with likelihood $\Pr(\hat{\theta}|\pi)$
- The null hypothesis is $H_0 : \pi \in \Omega_0$ and the alternative hypothesis is $H_1 : \pi \in \Omega_1$, for a partition $\Omega_0 \cup \Omega_1 = \Omega$
- Tests are mappings $\mathcal{T} : \mathbb{R}^K \to \{0, 1\}$, where $\mathcal{T} = 1$ corresponds to rejecting H_0, and for $S \subseteq \mathbb{R}^K$, let $\mathcal{T}(S) := I(\hat{\theta} \in S)$.
- The significance level of $\mathcal{T}(S)$ is $\sup_{\pi \in \Omega_0} Pr(\hat{\theta} \in S|\theta_1 \cdots \theta_K \sim \pi)$.
- The mean significance level of $\mathcal{T}(S)$ is $Pr(\hat{\theta} \in S|\pi \in \Omega_0)$.
- The mean power of $\mathcal{T}(S)$ is $Pr(\hat{\theta} \in S|\pi \in \Omega_1)$.

Mean most powerful tests - cont.

- MMP tests are Bayes rules for the following loss function:

$$L(S;\lambda_1,\lambda_2) = \lambda_1 \cdot I(\hat{\theta} \in S,\ \pi \in \Omega_0) + \lambda_2 \cdot I(\hat{\theta} \notin S,\ \pi \in \Omega_1).$$

- For the case that $\Omega_0 \cup \Omega_1 = \Omega$ the Bayes rule is

$$S^{Bayes}(\lambda_1,\lambda_2) = \{\hat{\theta} :\ \delta(\lambda_1,\lambda_2) \geq \Pr(\pi \in \Omega_0|\hat{\theta})\ \}$$

- NP type result: $\forall\delta$, $\mathcal{T}(S^{Bayes}(\delta))$ has greater or equal mean power than all tests with smaller or equal mean significance level
- We set threshold δ_α that controls significance level of the test

$$\sup_{\pi\in\Omega_0} Pr(\hat{\theta} \in S(\delta_\alpha)|\theta_1 \cdots \theta_K \sim \pi) \leq \alpha$$

A MMP test for distributions

To compute our statistic we can use *any* $\mathcal{D}$ that assigns probabilities to distributions. We use $\mathcal{D}$ defined as follows:

1. We partition $[a, b] \subseteq \mathbb{R}$ into I subintervals $[a_0, a_1] \cdots [a_{I-1}, a_I]$, with $A = \cup_{i \in I_A} [a_{i-1}, a_i]$ for $I_A \subset \{1 \cdots I\}$
2. We consider distributions that are step function in this partition

$$\pi(\theta) = \pi_1 \cdot \frac{I(\theta \in [a_0, a_1])}{a_1 - a_0} + \cdots + \pi_I \cdot \frac{I(\theta \in [a_{I-1}, a_I])}{a_I - a_{I-1}}$$

3. We use the one-to-one correspondence between $\pi(\theta)$ and $\vec{\pi} = (\pi_1 \cdots \pi_I)$ to define $\mathcal{D}(\pi)$ as the $Dirichlet(\vec{\pi}, \vec{\alpha})$ density

Thus $\Omega_0(\tilde{p}, A) = \{\pi : p_A \leq \tilde{p}\}$ can be expressed $\{\vec{\pi} : \pi_A \leq \tilde{p}\}$ for $\pi_A = \sum_{i \in I_A} \pi_i$ and the test statistic becomes:

$$\Pr(\pi \in \Omega_0 | \hat{\theta}) = \Pr(\pi_A \leq \tilde{p} | \hat{\theta})$$

The MMP test in the no noise case

- For $n_i = \#\{k : \hat{\theta}_k \in [a_{i-1}, a_i]\}$,

$$\vec{\pi} | \, \vec{n} \sim \mathit{Dirichlet}(\vec{\alpha} + \vec{n})$$

- Thus for the corresponding sums over A and $A^C = [a,b] - A$,

$$\pi_A | \, \vec{n} \sim \mathit{Beta}(\alpha_A + n_A, \alpha_{AC} + n_{AC})$$

This means that:

1. Our statistic is very easy to compute

$$\Pr(\pi_A \leq \tilde{p} | \hat{\theta}) = \mathit{pbeta}(\tilde{p}, \alpha_A + n_A, \alpha_{AC} + n_{AC})$$

2. For any choice of intervals and any choice of $\vec{\alpha}$ the MMP test sorts the data sample space according to n_A (= naive test)!!!!!!!!

The MMP test in the general noisy case

Our model assumes the data is generated hierarchically:

1. Generate $\vec{\pi} \sim \mathit{Dirichlet}(\vec{\alpha})$
2. Generate iid $\delta_1 \cdots \delta_K$, $\Pr(\delta_k = i) = \pi_i$ for $i = 1 \cdots I$
3. For $\delta_k = i$, generate $\theta_k \sim U[a_{i-1}, a_i]$
4. For θ_k, generate $\hat{\theta}_k \sim f(\hat{\theta}_k | \theta_k)$

Expressing the conditional distribution of π given $\hat{\theta}_1 \cdots \hat{\theta}_K$,

$$\begin{aligned}
\Pr(\pi|\hat{\theta}) &= \frac{\Pr(\pi, \hat{\theta})}{\Pr(\hat{\theta})} = \frac{\sum_{\vec{n}} \Pr(\pi, \hat{\theta}, \vec{\delta} = \vec{n})}{\sum_{\vec{n}} \Pr(\hat{\theta}, \vec{\delta} = \vec{n})} \\
&= \frac{\sum_{\vec{n}} \Pr(\pi|\hat{\theta}, \vec{\delta} = \vec{n}) \cdot \Pr(\hat{\theta}, \vec{\delta} = \vec{n})}{\sum_{\vec{n}} \Pr(\hat{\theta}|\vec{\delta} = \vec{n}) \cdot \Pr(\vec{\delta} = \vec{n})} \\
&= \frac{\sum_{\vec{n}} \Pr(\pi|\vec{\delta} = \vec{n}) \cdot \Pr(\hat{\theta}|\vec{\delta} = \vec{n}) \cdot \Pr(\vec{\delta} = \vec{n})}{\sum_{\vec{n}} \Pr(\hat{\theta}|\vec{\delta} = \vec{n}) \cdot \Pr(\vec{\delta} = \vec{n})}.
\end{aligned}$$

$\Rightarrow \pi|\hat{\theta}$ is a mixture of Dirichlet distributions

Computing the test statistic

To compute our test statistic we express

$$\Pr(\pi|\hat{\theta}) = \frac{\Pr(\pi,\hat{\theta})}{\Pr(\hat{\theta})} = \frac{\Pr(\hat{\theta}|\pi)\cdot\Pr(\pi)}{\Pr(\hat{\theta})} = \frac{[\Pi_{k=1}^{K} f(\hat{\theta}_k|\pi)]\cdot\Pr(\pi)}{\Pr(\hat{\theta})}$$

where we numerically approximate

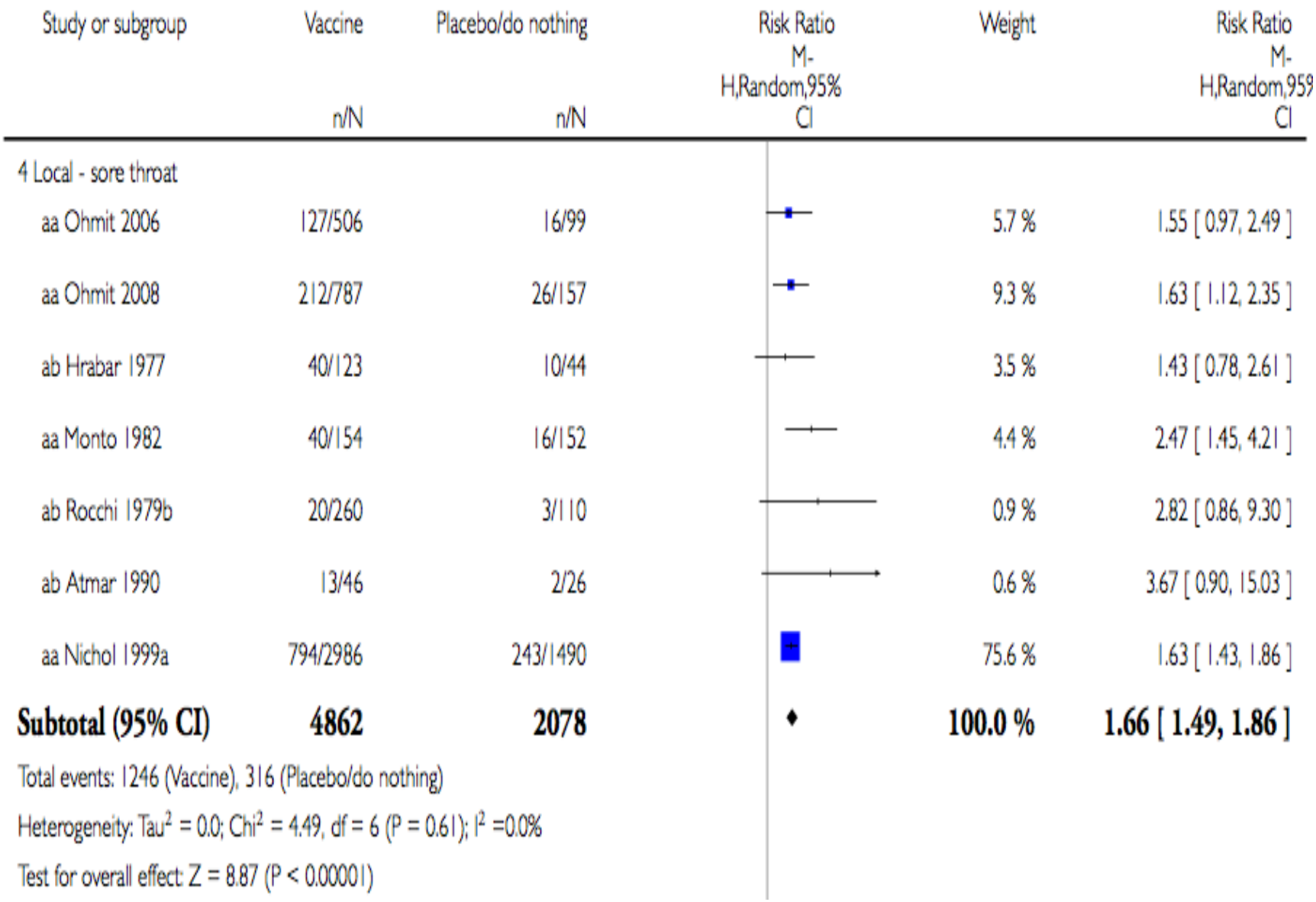
$$f(\hat{\theta}_k|\pi) = \int_{\theta_k} f(\hat{\theta}_k|\ \theta_k)\pi(\theta_k)d\theta$$

Algorithm:

1. Draw 10^5 iid realiztions $\pi^l \sim Dirichlet(\alpha)$
2. For $l = 1 \cdots 10^5$, compute $\Pr(\hat{\theta}|\pi^l)$ and π_A^l
3. Compute

$$\Pr(\pi_A \leq \tilde{p}|\hat{\theta}) \ = \ \sum_{\pi^l \leq \tilde{p}} \Pr(\hat{\theta}|\pi^l) / \sum_{l=1}^{10^5} \Pr(\hat{\theta}|\pi^l).$$

Example 1: 95% CI for $p_{[0,\infty]}$ for the sore throat outcome



Study or subgroup	Vaccine n/N	Placebo/do nothing n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
4 Local - sore throat					
aa Ohmit 2006	127/506	16/99		5.7 %	1.55 [0.97, 2.49]
aa Ohmit 2008	212/787	26/157		9.3 %	1.63 [1.12, 2.35]
ab Hrabar 1977	40/123	10/44		3.5 %	1.43 [0.78, 2.61]
aa Monto 1982	40/154	16/152		4.4 %	2.47 [1.45, 4.21]
ab Rocchi 1979b	20/260	3/110		0.9 %	2.82 [0.86, 9.30]
ab Atmar 1990	13/46	2/26		0.6 %	3.67 [0.90, 15.03]
aa Nichol 1999a	794/2986	243/1490		75.6 %	1.63 [1.43, 1.86]
Subtotal (95% CI)	**4862**	**2078**		**100.0 %**	**1.66 [1.49, 1.86]**

Total events: 1246 (Vaccine), 316 (Placebo/do nothing)

Heterogeneity: $Tau^2 = 0.0$; $Chi^2 = 4.49$, df = 6 (P = 0.61); $I^2 = 0.0\%$

Test for overall effect: Z = 8.87 (P < 0.00001)

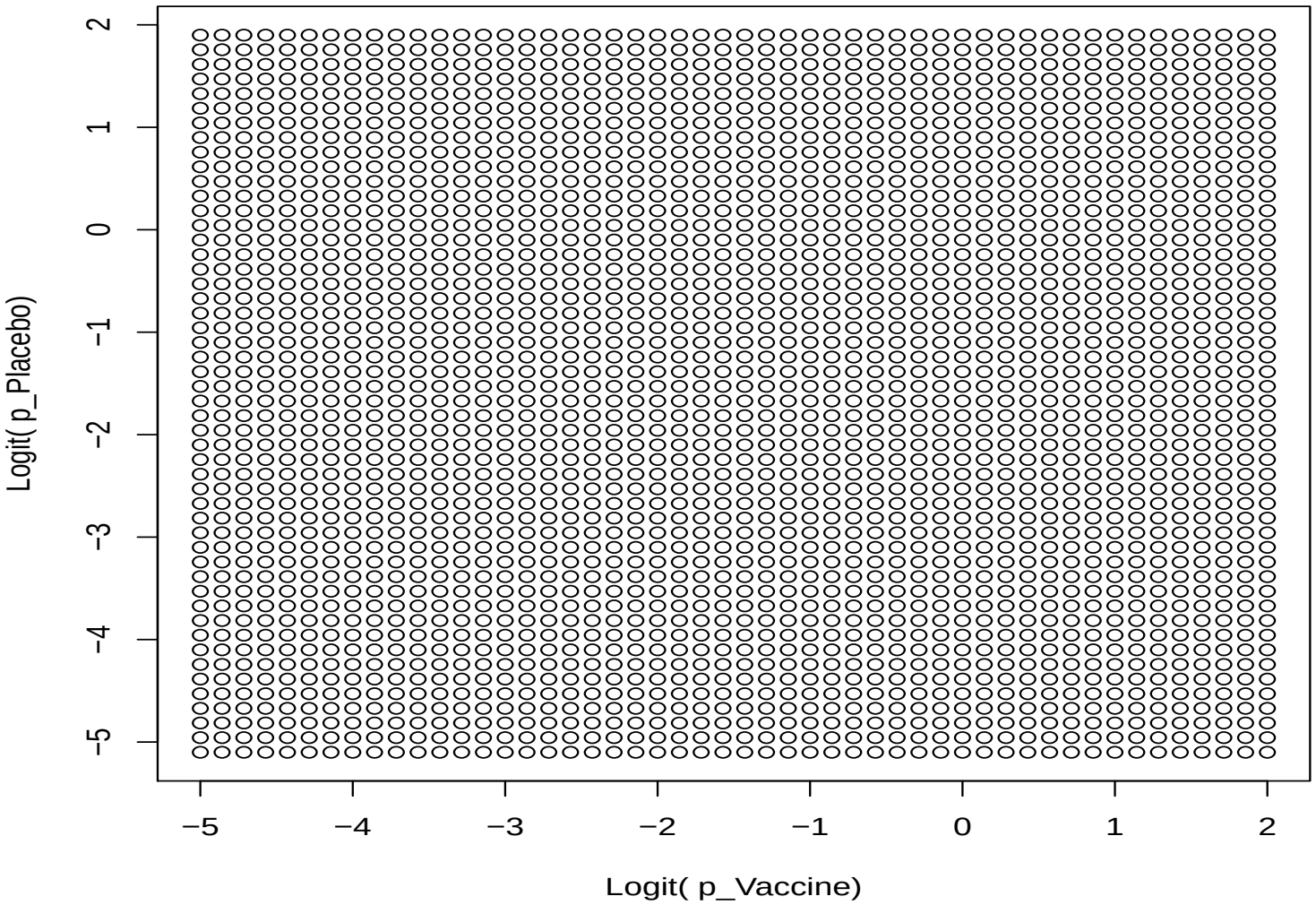
Parameterization for sore throat outcome review

- Studies $k = 1 \cdots 7$
- Data: # of Vaccine treated $n_{k,V}$, # of Placebo treated $n_{k,P}$, # of Vaccine affected $X_{k,V}$, # of Placebo affected $X_{k,P}$
- Parameters: $\theta_k = (p_V, p_P)$, for Vaccine risk p_V and Placebo risk p_P
- Likelihood:

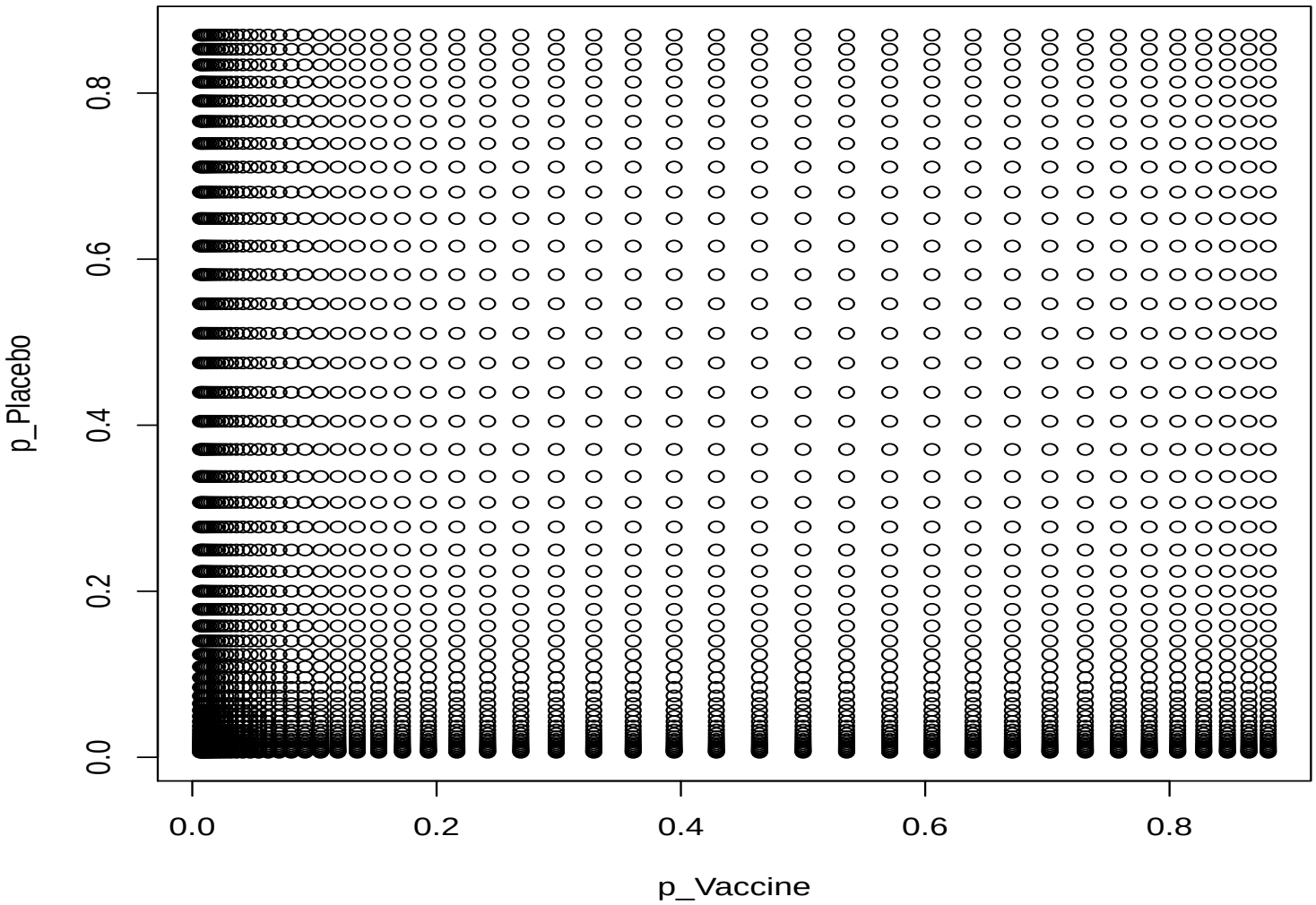
$$X_{k,V} \sim Binom(n_{k,V},\ p_{k,V}) \text{ and } X_{k,P} \sim Binom(n_{k,P},\ p_{k,P})$$

- The π's are step function on 6 "interval" partition of $[0,1] \times [0,1]$ space of $(p_{k,V}, p_{k,P})$ with $\vec{\alpha} = (1.676, 0.564, 0.210, 0.210, 0.564, 1.776)$
- $\Omega_0(\tilde{p}, A) = \{\pi : p_A \leq \tilde{p}\}$ is the set of distributions that give probability less than $\tilde{p}$ to the event $p_V \geq p_P$ (or $log(RR) \in [0, \infty]$)

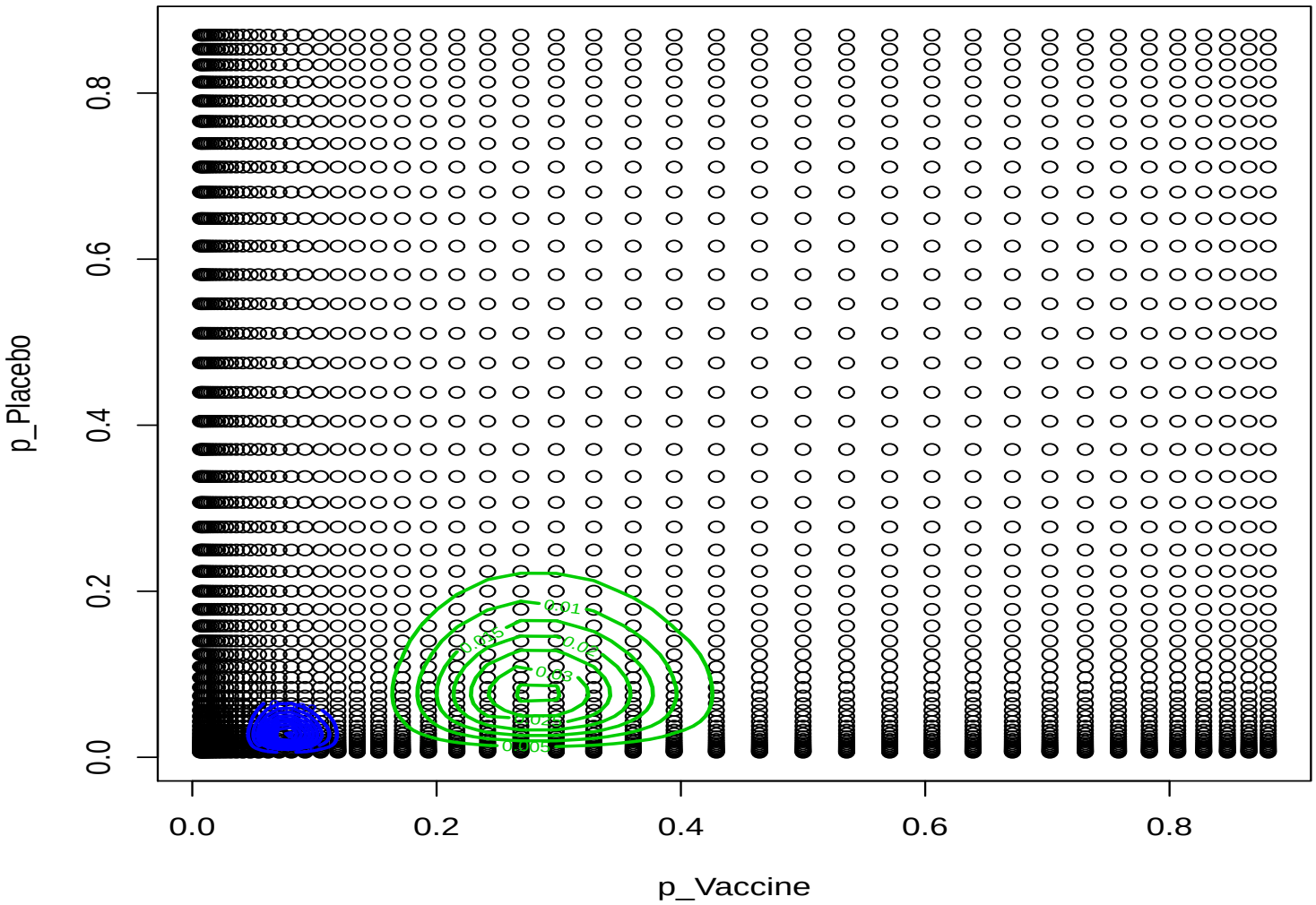
Support of π Logit scale



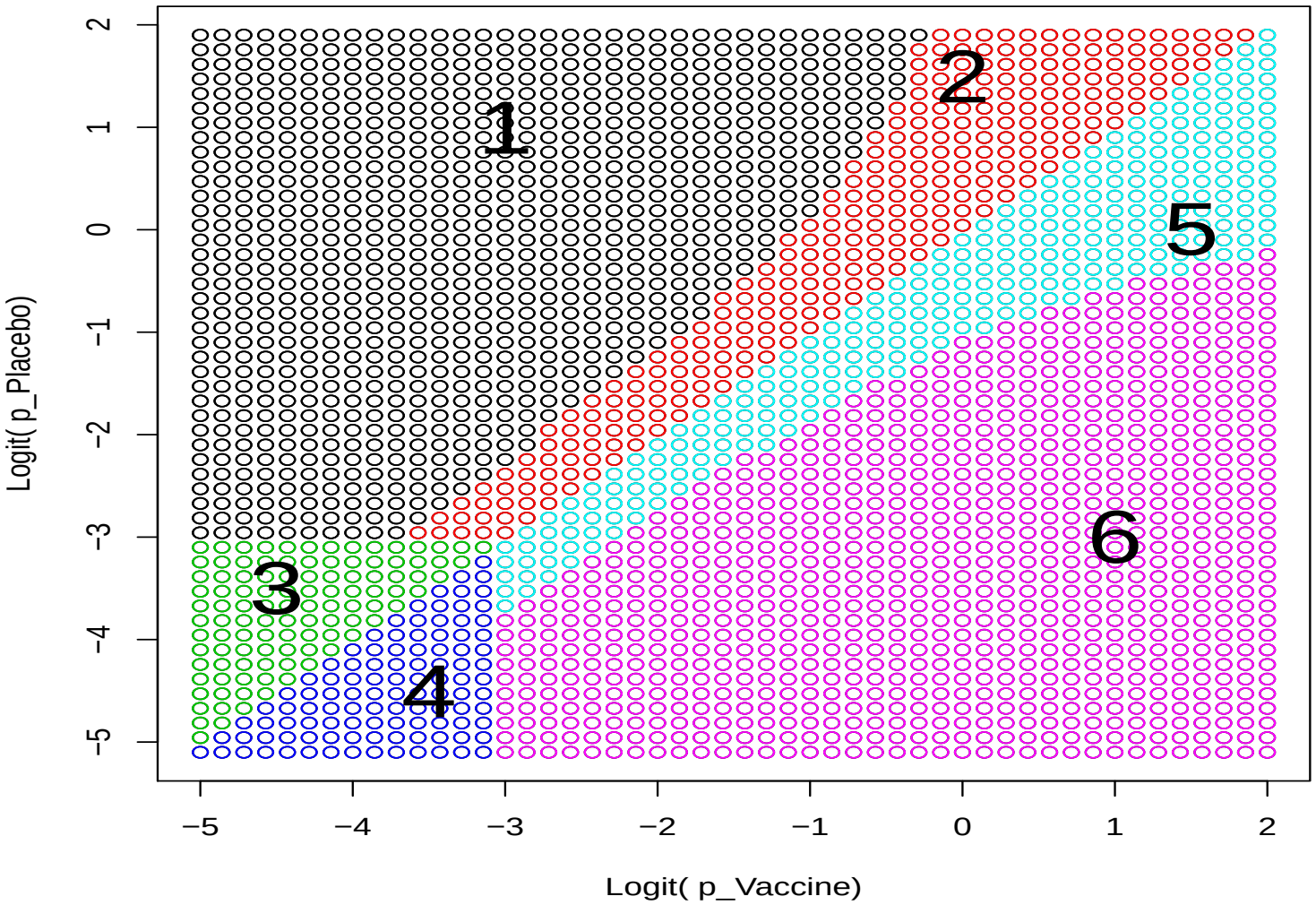
Support of π



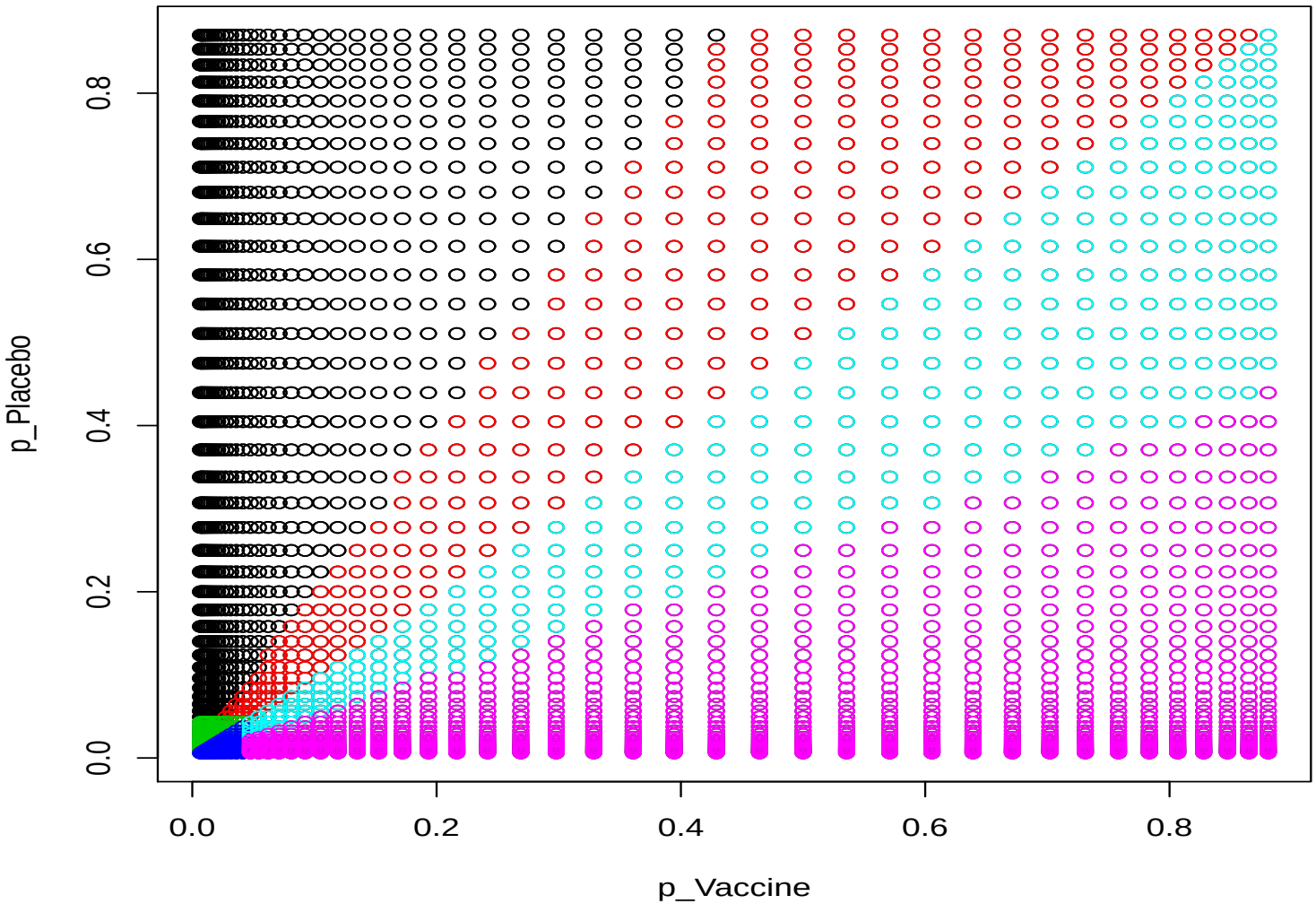
(p_V, p_P) Likelihood for Studies 5 (blue) and 6 (green)



π intervals in Logit scale: $\Omega_0 = Int_1 \cup Int_2 \cup Int3$



π intervals in p scale



Let's begin testing!

Warm up: $H_0 : p_{[0,\infty]} < 0.5$

- $\Pr(p_{[0,\infty]} < 0.5 |$ observed data$) = 0.0150$
- Is this a small value?
  $\Pr(p_{[0,\infty]} < 0.5 |$ worst data$) = 0.0135$ while
  $1 - pbeta(.5, \alpha_1 \cdots \alpha_3, \alpha_4 \cdots \alpha_6 + 7) = 0.0014$
- Is this significant?
  We computed statistic for 10^3 data realizations from worst (?) null π (takes about a minute) and we got smaller test statistic values in $7/1000$ realizations.

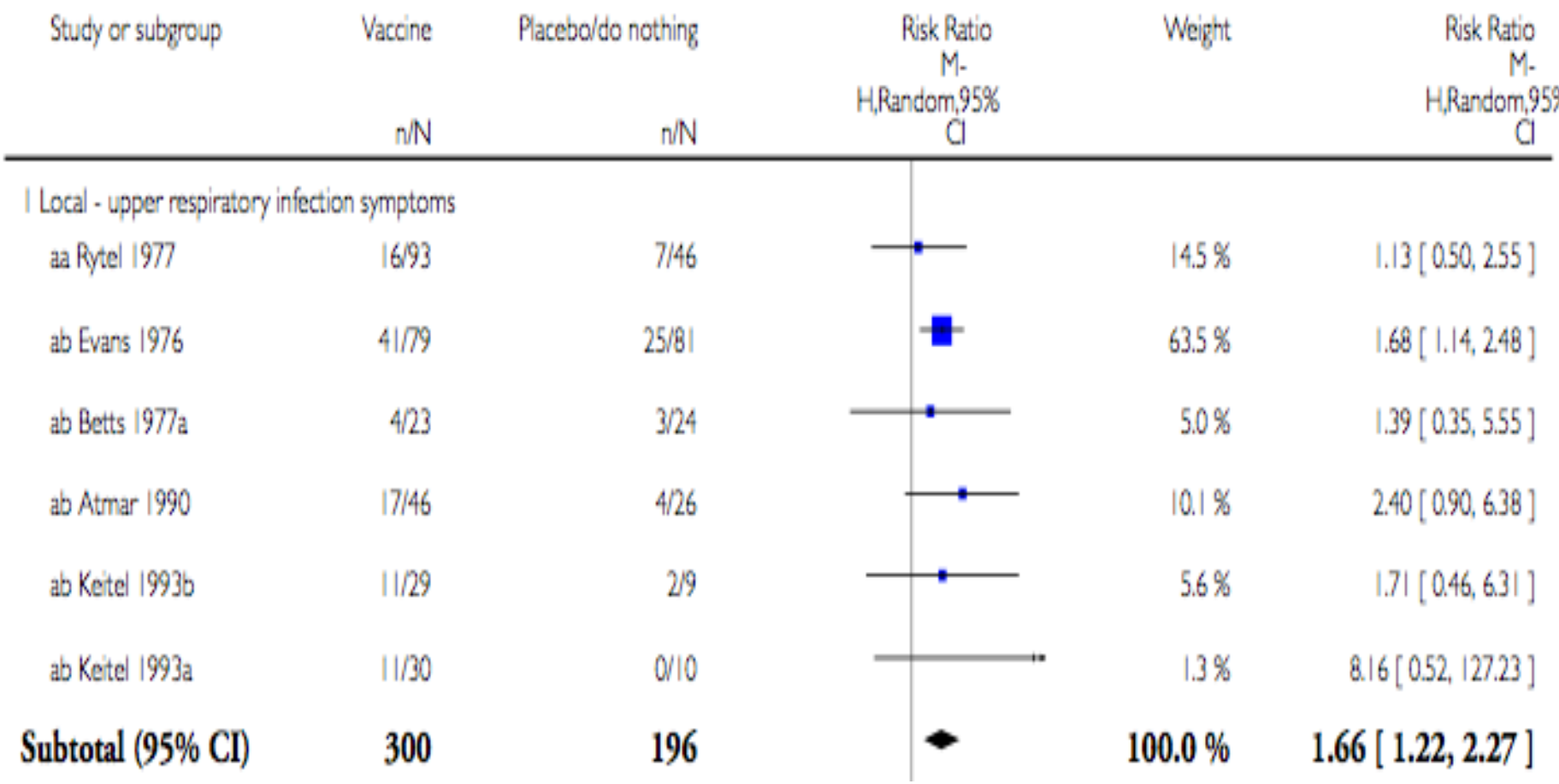
Continue testing

$> dbinom(7, 7, c(.5, .6, .65, .66, .7)) = 0.0080.0280.0490.0550.082$

- We set $\tilde{p} = .55, .60, .62, .61 \cdots .70$ and compared $\Pr(p_{[0,\infty]} < \tilde{p} |$ observed data$)$ with distribution of $\Pr(p_{[0,\infty]} < \tilde{p} |$ data$)$ for 10^3 data realizations from corresponding worst (?) null π.
- Results: H_0 is accepted for $0.64 \leq \tilde{p}$

$\Rightarrow$ 95% CI for $p_{[0,\infty]}$ is $[0.64, 1]$

Example 2: Upper respiratory infection symptoms outcome



Study or subgroup	Vaccine n/N	Placebo/do nothing n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
1 Local - upper respiratory infection symptoms					
aa Rytel 1977	16/93	7/46		14.5 %	1.13 [0.50, 2.55]
ab Evans 1976	41/79	25/81		63.5 %	1.68 [1.14, 2.48]
ab Betts 1977a	4/23	3/24		5.0 %	1.39 [0.35, 5.55]
ab Atmar 1990	17/46	4/26		10.1 %	2.40 [0.90, 6.38]
ab Keitel 1993b	11/29	2/9		5.6 %	1.71 [0.46, 6.31]
ab Keitel 1993a	11/30	0/10		1.3 %	8.16 [0.52, 127.23]
Subtotal (95% CI)	**300**	**196**		**100.0 %**	**1.66 [1.22, 2.27]**

Let's test!

> dbinom(6,6,c(.6,.61,.62)) [1] 0.04665600, 0.05152037, 0.05680024

or maybe

> dbinom(6,6,c(.59,.60,.61)) [1] 0.04218053, 0.04665600, 0.05152037

- We set $\tilde{p} = .30, .35, \cdots, .60$ and compared $\Pr(p_{[0,\infty]} < \tilde{p} | \text{ observed data})$ with distribution of $\Pr(p_{[0,\infty]} < \tilde{p} | \text{ data})$ for 10^3 data realizations from corresponding worst (?) null π.
- Results: observed statistic is smaller than null samples 0.05 quantile for $\tilde{p} = 0.35$ and larger than null samples 0.05 quantile for $0.40 \leq \tilde{p}$

$\Rightarrow$ 95% CI for $p_{[0,\infty]}$ is $[0.40, 1]$

Work in progress!!

- Improve numerics
- Determining worst null π
- Confidence statements for conditional inference