Introduction to Bayesian Meta-analysis

Hwanhee Hong, PhD
Associate Professor
Department of Biostatisitcs and Bioinformatics
Duke School of Medicine

March 27, 2025
Guest Lecture
BIOS 725 Bayesian Health Data Science
(Instructor: Dr. Sam Berchuck)
Duke University

Outline

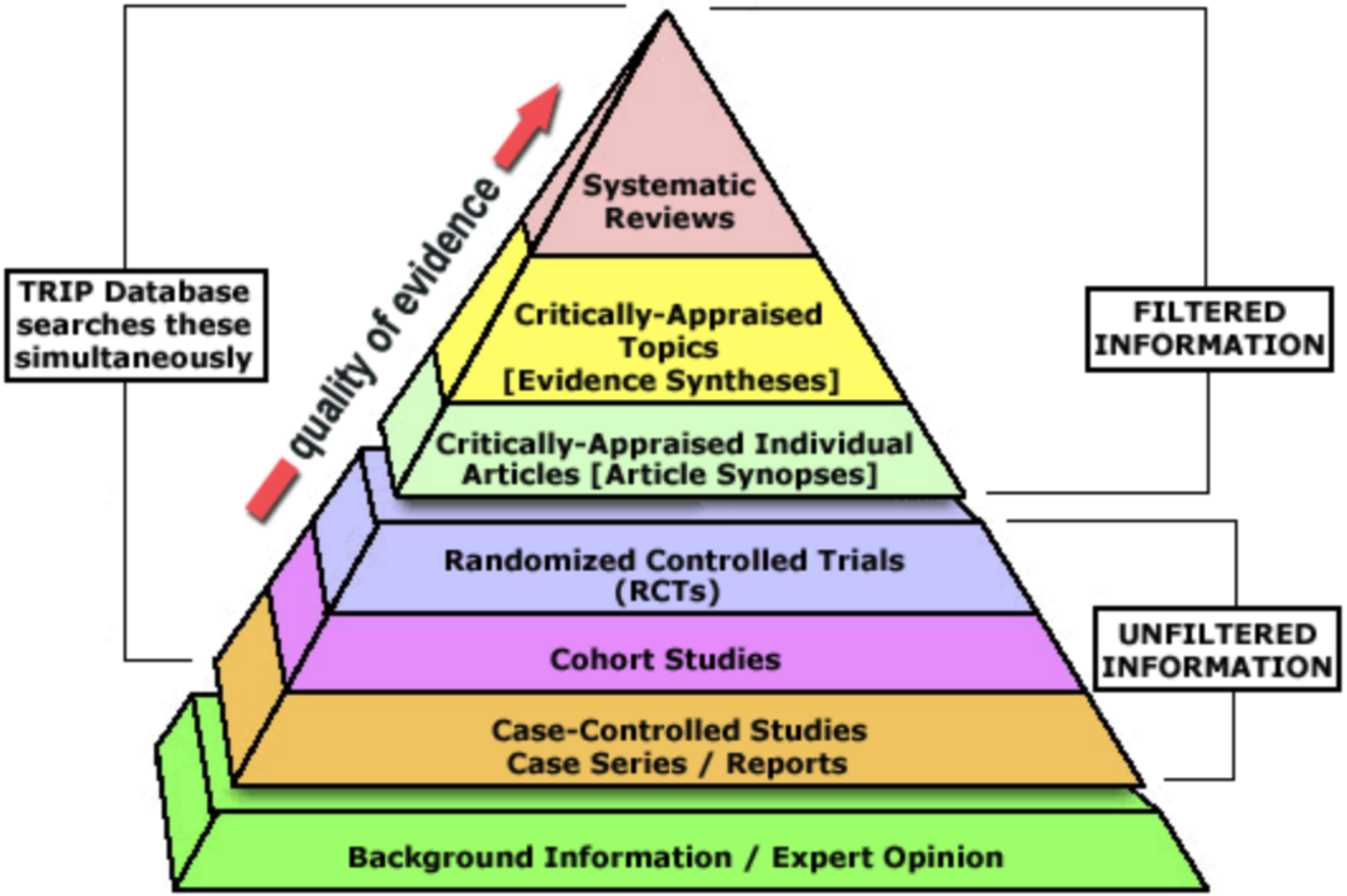
Outline

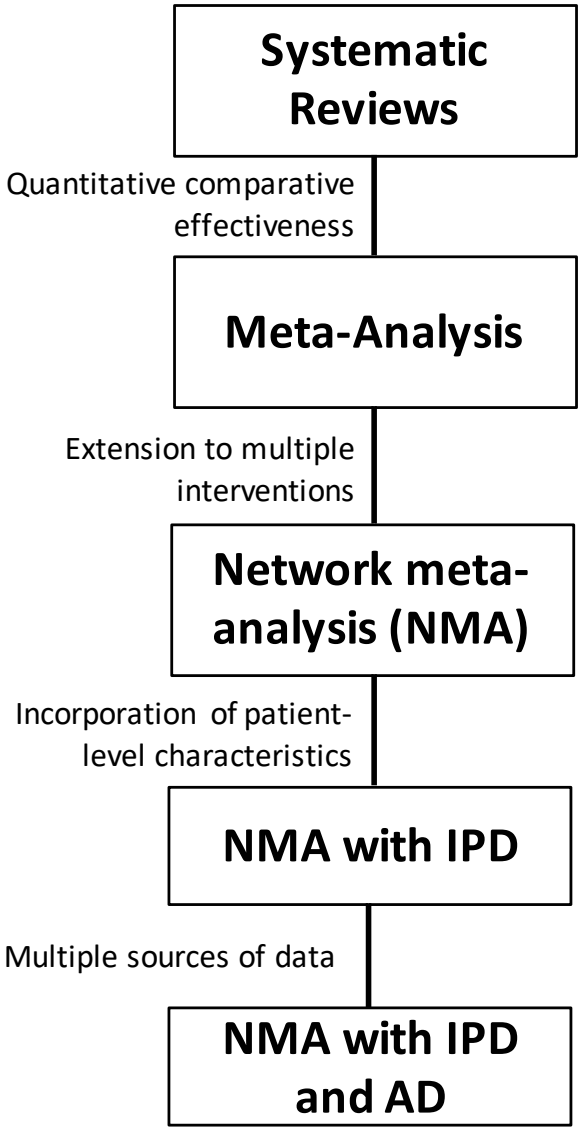
Comparative effectiveness research

- Comparative effectiveness research (CER) helps answer "what works best" and inform health-care decisions by providing evidence on the effectiveness, benefits, and harms of different treatments.[1]
- The evidence can be obtained from existing or new research studies that compare drugs, medical devices, or ways to deliver health care.
- The goal can be achieved by synthesizing all relevant study findings in a principled way about key aspects of efficacy or safety of certain interventions.
- CER also helps improve evidence-based medicine, personalized medicine, and the use of current best evidence in making decisions about the care of individual patients.

[1] Agency for Healthcare Research and Quality (AHRQ) http://www.ahrq.gov

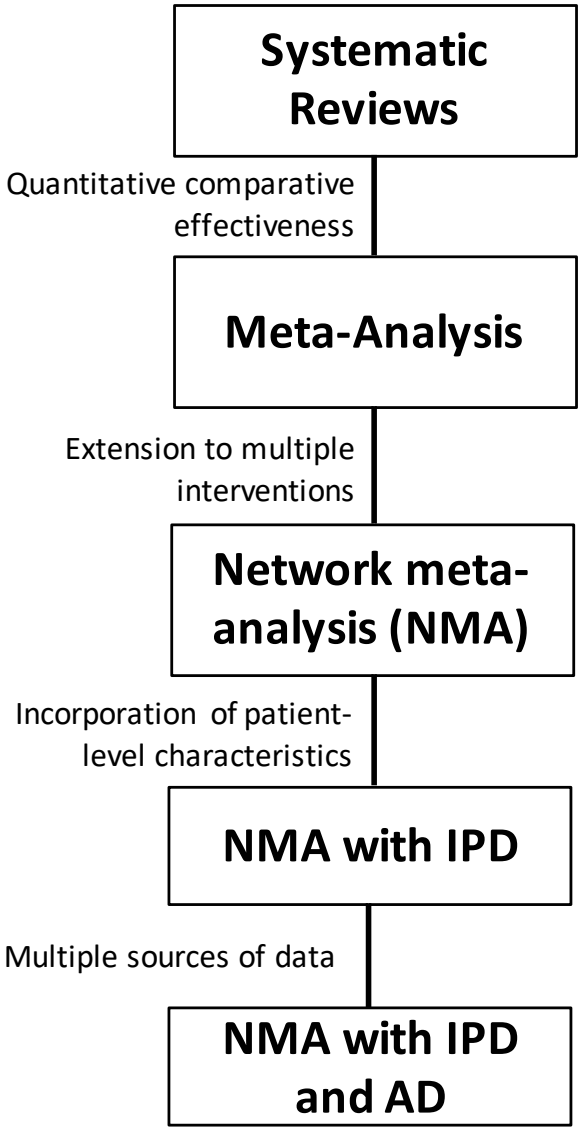
Hierarchy of evidence





- Integration and collection of both published and unpublished findings for a specific question
- Comparison of common intervention and comparator
- Combine direct and indirect evidence
- Key issues: heterogeneity and inconsistency
- Suitable analysis for patient-specific decision-making
- Adaptive borrowing between different data source

IPD: individual participant-level data
AD: aggregate-level data



- Frequentist vs. Bayesian
- Rare event

- Frequentist vs. Bayesian
- Contrast-based vs. arm-based models
- Multiple outcomes
- Single-arm trials

- Multiple outcomes
- Generalizability
- Causal inference
- Missing data
- Treatment effect heterogeneity

- Power and commensurate priors
- Bayesian models

IPD: individual participant-level data
AD: aggregate-level data

Outline

Frequentist vs. Bayesian meta-analysis

Article

CLINICAL TRIALS

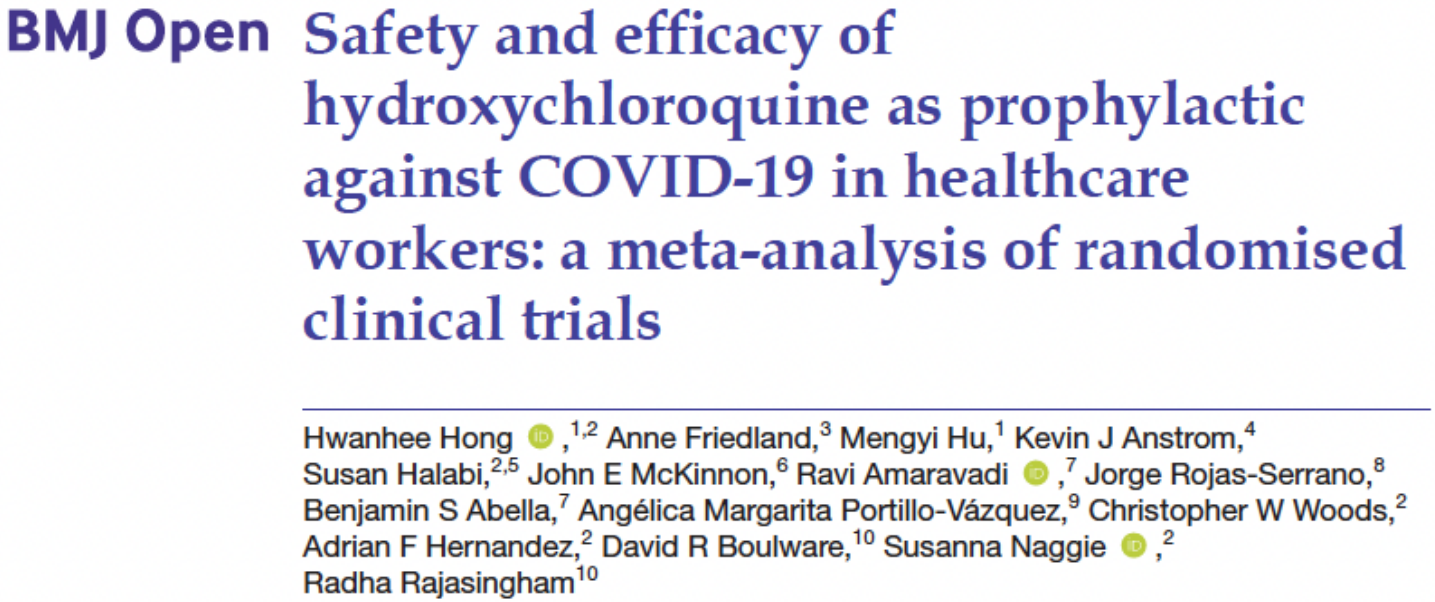
Clinical Trials
1–14
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1740774520969136
journals.sagepub.com/home/ctj
SAGE

Meta-analysis of rare adverse events in randomized clinical trials: Bayesian and frequentist methods

Hwanhee Hong[1], Chenguang Wang[2] and Gary L Rosner[2]

Hong et al. (2020). *Clinical Trials* 18(1):1-14

An example of meta-analysis



BMJ Open

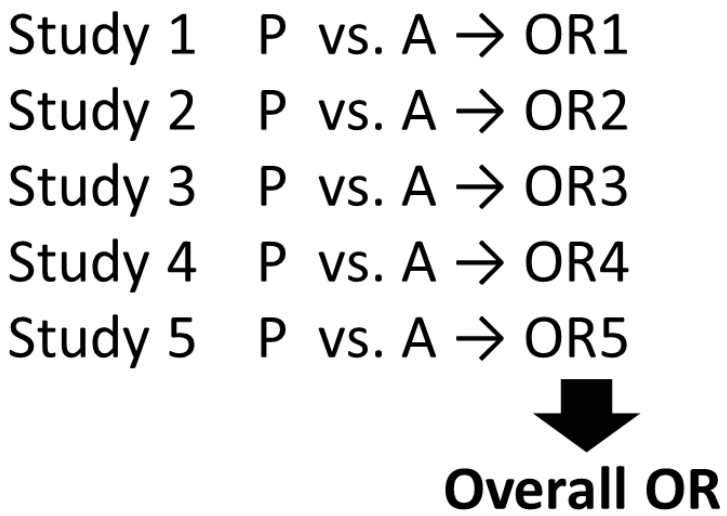
Safety and efficacy of hydroxychloroquine as prophylactic against COVID-19 in healthcare workers: a meta-analysis of randomised clinical trials

Hwanhee Hong ,[1,2] Anne Friedland,[3] Mengyi Hu,[1] Kevin J Anstrom,[4] Susan Halabi,[2,5] John E McKinnon,[6] Ravi Amaravadi ,[7] Jorge Rojas-Serrano,[8] Benjamin S Abella,[7] Angélica Margarita Portillo-Vázquez,[9] Christopher W Woods,[2] Adrian F Hernandez,[2] David R Boulware,[10] Susanna Naggie ,[2] Radha Rajasingham[10]

Hong et al. (2023). *BMJ Open* 13:e065305

Meta-analysis

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



Set up the research question

To address the most common limitation, inadequate power to show a modest effect, we conducted a formal meta-analysis of pre-exposure prophylactic HCQ studies in HCWs. We conducted a systematic search for clinical trials of pre-exposure prophylactic use of HCQ against infection of SARS-CoV-2 in HCWs, thoroughly compared similarities and differences in characteristics of the identified studies and performed a Bayesian meta-analysis to combine results of the trials.

Hong et al. (2023). *BMJ Open* 13:e065305

Eligibility criteria (population)

Eligibility criteria and study selection

The eligibility criteria included phase II or phase III RCTs of HCQ for use as PrEP in HCWs with moderate to high risk of exposure. We excluded observational studies, crossover trials, studies where the method of allocation to treatment was not truly random, duplicate studies and non-original data studies. No language, publication date or publication status restrictions were applied. References of prior systematic reviews and meta-analyses were also screened for related studies. Study selection involved screening of titles and abstracts followed by full-text evaluation of possible eligible studies.

Hong et al. (2023). *BMJ Open* 13:e065305

Search strategy

Search strategy and information sources

We searched PubMed/Medline and Ovid/EMBASE databases from database inception through the final search date, 14 March 2023. We used keywords related to COVID-19, HCQ and RCTs. The full search strategies are provided in online supplemental table 1.

eTable 1. Search code that was used to identify publications as of March 14, 2023

PubMed search

#1	covid[Title] OR coronavirus[Title] OR sars-cov-2[Title]
#2	hydroxychloroquine[Title]
#3	randomized[Title/Abstract] OR randomized[Title/Abstract]
#4	#1 AND #2 AND #3

Embase search

#1	covid:ti OR coronavirus:ti OR 'sars cov 2':ti
#2	hydroxychloroquine:ti
#3	randomized:ab,ti OR randomised:ab,ti
#4	#1 AND #2 AND #3

Flow chart

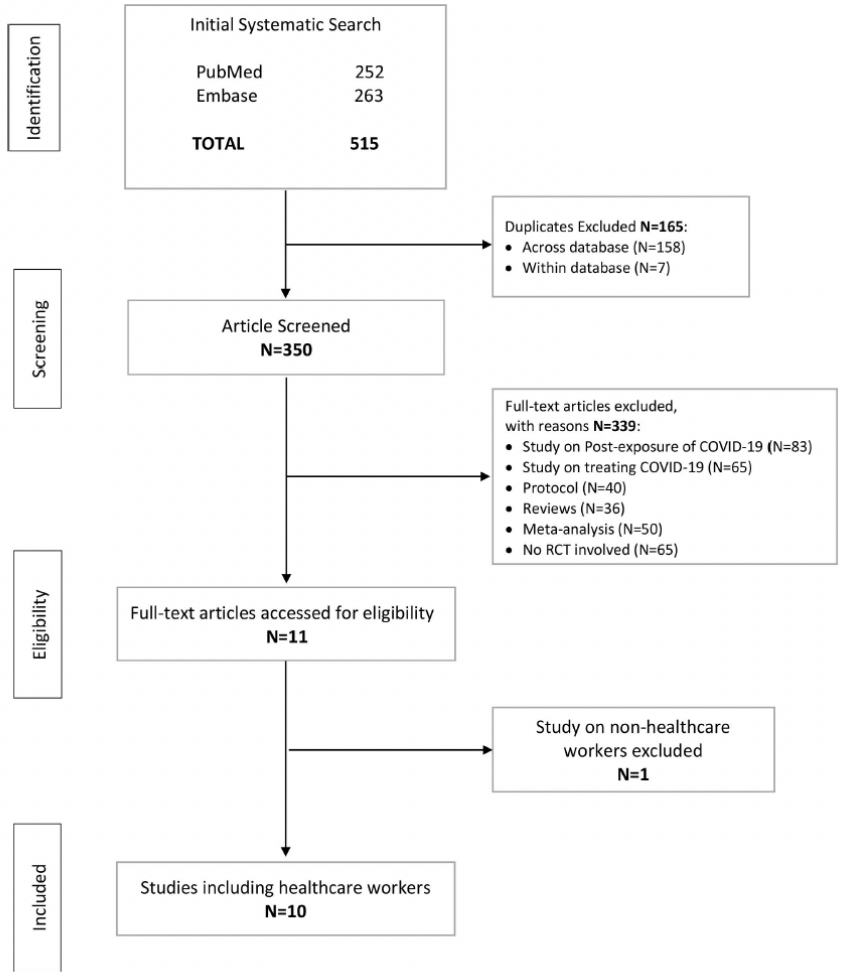


Figure 1 Flow chart of literature review. RCT, randomised controlled trial.

Hong et al. (2023). *BMJ Open* 13:e065305

Data collection process

Data collection process

Each of the selected studies was independently reviewed by two reviewers (AF, MH or HH). We extracted data on the study design, baseline characteristics, interventions and outcomes. Any disagreements of collected information between reviews were reconciled through discussion by all three reviewers.

Hong et al. (2023). *BMJ Open* 13:e065305

Outcome definition

Outcome measures

The primary efficacy outcome for the meta-analysis was laboratory-confirmed SARS-CoV-2 infection by PCR test and the primary safety outcome was incidence of adverse events (table 1). The secondary efficacy outcome was suspected or probable SARS-CoV-2 infection. Included studies had the following outcome definitions: (1) laboratory-confirmed SARS-CoV-2 infection defined as COVID-19-like symptoms and positive SARS-CoV-2 PCR and (2) suspected or probable SARS-CoV-2 infection defined as COVID-19-like symptoms but lack of confirmatory PCR testing.

Outcomes should be pre-specified.

Hong et al. (2023). *BMJ Open* 13:e065305

Meta-analysis outcome data

eTable 6. Results of outcome measures in trials included in the meta-analysis. Sample size and the number of participants who had each outcome are reported with proportions (%) in parentheses.

	Treatment	N (ITT)	Confirmed COVID-19	Suspected with COVID compatible symptoms	Adverse event[2]
Naggie et al. (HERO-HCQ)	HCQ	683	3 (0.4)	38 (5.6)	16 (2.3)
	Placebo	676	6 (0.9)	47 (7.0)	13 (1.9)
Abella et al. (PATCH)	HCQ	64	4 (6.3)		29 (45.3)
	Placebo	61	4 (6.6)		17 (27.9)
Rajasingham et al. (MN-COVID-PREP)	HCQ[1]	989	11 (1.1)	47 (4.8)	316 (32.0)
	Placebo	494	6 (1.2)	33 (6.7)	100 (20.2)
Rojas-Serrano et al.	HCQ	62	1 (1.6)		32 (51.6)
	Placebo	65	6 (9.2)		38 (58.5)
McKinnon et al. (WHIP)	HCQ[1]	387	2 (0.5)	22 (5.7)	192 (49.6)
	Placebo	191	2 (1.0)	13 (6.8)	85 (44.5)
Vijayaraghavan et al.	HCQ	211	11 (5.2)	12 (5.7)	21 (10.0)
	Placebo	203	12 (5.9)	12 (5.9)	14 (6.9)
Polo et al. (EPICOS)	HCQ	224	21 (9.4)		100 (44.6)
	Placebo	211	23 (10.9)		94 (44.5)
Llanos-Cuentas et al.	HCQ	34	5 (14.7)		
	Placebo	31	3 (9.7)		
Grau-Pujol et al.	HCQ	137	1 (0.7)	3 (2.2)	53 (38.7)
	Placebo	116	1 (0.9)	3 (2.6)	42 (36.2)
Syed et al.	HCQ[1]	154	42 (27.3)		9 (5.8)
	Placebo	46	7 (15.2)		1 (2.2)

HCQ= Hydroxychloroquine; ITT= Intention-to-treat; AE=Adverse event ; COVID-RS=COVID-19 related symptoms ; Vit C= Vitamin C

[1] More than one HCQ groups with different doses are lumped.

[2] Number of patients with any adverse events

Hong et al. (2023). *BMJ Open* 13:e065305

Risk of Bias

eTable 2. Risk of bias for trials included in the meta-analysis using the Cochrane risk assessment tool. Green circle is for low risk and yellow circle is for some concerns

	Selection bias (Randomization process)	Performance bias (Deviations from the intended interventions)	Attrition bias[1] (Missing outcome data)	Reporting bias (Measurement of the outcome)	Other sources of bias (Selection of the reported result)
Naggie et al. (HERO-HCQ)					
Abella et al. (PATCH)					
Rajasingham et al. (MN-COVID-PREP)					
Rojas-Serrano et al.					
McKinnon et al. (WHIP)					
Vijayaraghavan et al.					
Polo et al. (EPICOS)					
Llanos-Cuentas et al.					
Grau-Pujol et al.					
Syed et al.					

[1] The Rojas-Serrano et al. study reported minimal loss to follow-up (<10%). The Rojas-Serrano et al. study reported 18% (25/130) lost to follow-up and additional 12% (16/130) discontinued the intervention.

Hong et al. (2023). *BMJ Open* 13:e065305

Meta-analysis: Data structure

1. Continuous outcome
   Sample mean, sample size, and sample standard deviation

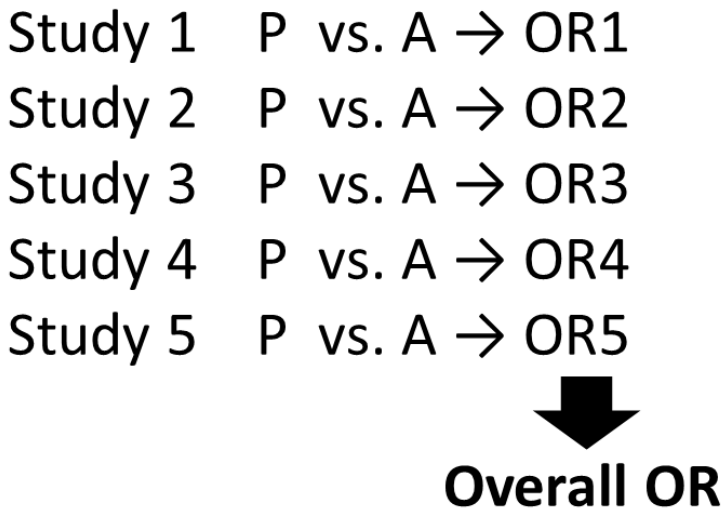
	Control	Treated
Study 1	6.9 (2.0)	3.8 (2.2)
Study 2	4.1 (1.6)	3.7 (2.3)
Study 3	2.5 (2.5)	2.3 (2.8)

2. Binary outcome

	Control	Treated
Study 1	25/100	48/150
Study 2	10/80	23/100
Study 3	14/130	30/150

Meta-analysis

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



- **Fixed effect** or **common treatment effect (CTE)** model assumes a common treatment effect across studies. E.g., $LOR_i \equiv LOR$
- **Random effects** or **heterogeneous treatment effect (HTE)** model assumes that treatment effects vary across studies. E.g., $LOR_i \sim N(LOR, \tau^2)$

Fixed vs. random effects meta-analysis

- **Fixed effect** MA assumes that all studies estimate the same effect size, and therefore there is no variability between studies.
- The inverse-variance weight in a fixed effect MA is

$$w_{f,i} = \frac{1}{\sigma_i^2/n_i},$$

where σ_i^2/n_i measures the within-study effect variability in the i^{th} study.

- **Random effects** MA assumes that the true effects in the studies are sampled from a distribution of true effects, and allows heterogeneity between studies.
- The inverse-variance weight in a random effect MA is

$$w_{r,i} = \frac{1}{\sigma_i^2/n_i + \tau^2},$$

where τ^2 is the between-study effect variability.

Fixed vs. random effects meta-analysis

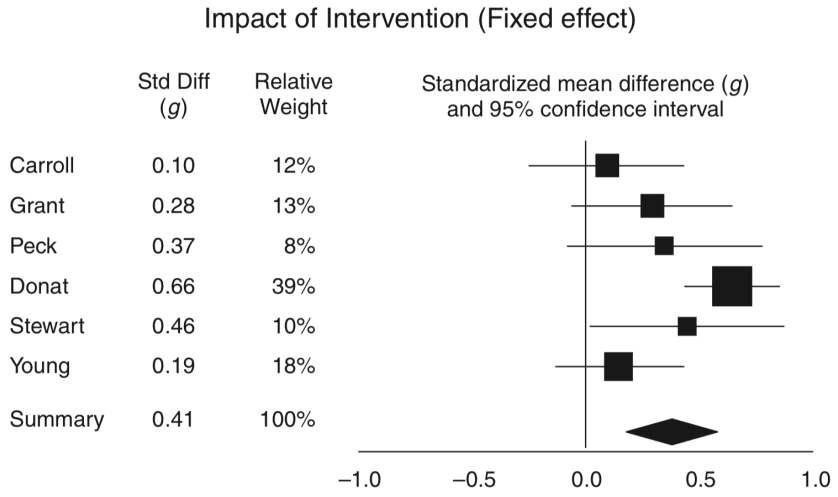


Figure 13.1 Fixed-effect model – forest plot showing relative weights.

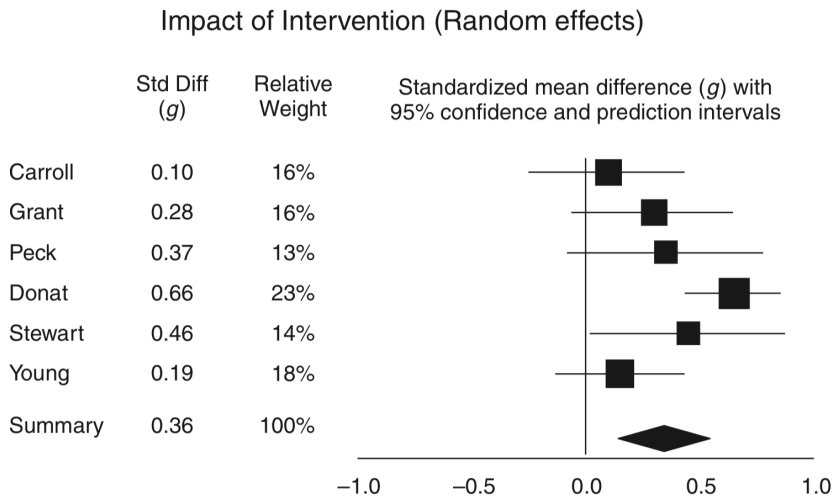


Figure 13.2 Random-effects model – forest plot showing relative weights.

Introduction to Meta-Analysis. Borenstein et al. 2009

- Study weights are more balanced under the random effects MA.
- Standard error of the summary effect is larger under the random effects MA.

Report heterogeneity

- Heterogeneity refers to treatment effect size variability between studies (possibly due to design and patient characteristics).
- Heterogeneity can be measured by I^2 or the estimated variance of random effects, τ^2, by fitting a random effect model.
- I^2 presents the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error.
- I^2 less than 40% might not be important, 30% to 60% may represent moderate heterogeneity, and larger than 60% represents substantial heterogeneity.
- Do not solely rely on values. The actual data should be investigated.

Frequentist approaches

- Fixed effect MA: Peto and Mentel Haenszel (MH).
- Random effects MA: DerSimonian and Laird Inverse-Variance (IV)
- For rare binary outcomes, data modification (or continuity correction) is applied to for studies having one or more zero cells.

	No. of events	No. of nonevents	Total
Control group	$y_{i1} + 0.5$	$n_{i1} - y_{i1} + 0.5$	$n_{i1} + 1$
Treated group	$y_{i2} + 0.5$	$n_{i2} - y_{i2} + 0.5$	$n_{i2} + 1$
Total	$y_{i\cdot} + 1$	$n_{i\cdot} - y_{i\cdot} + 1$	$n_{i\cdot} + 2$

Table: DM using a constant

- `metafor` in R, `metan` in Stata

Hong et al. (2020) *Clinical Trials*. Meta-analysis of rare adverse events in randomized clinical trials: Bayesian and frequentist methods

Bayesian approach: logistic regression

$$\begin{aligned} \text{Fixed effect MA: } logit(p_{ik}) &= \mu_i + d * I(k = 2) \\ \text{Random effects MA: } logit(p_{ik}) &= \mu_i + \delta_i * I(k = 2) \end{aligned}$$

- μ_i is the study-specific baseline effect (i.e., log odds of control group for study i).
- d and δ_i are the log odds ratio (LOR) between treated and control groups.

- We distinguish the two models based on the assumption that whether we allow heterogeneity of treatment effects via LOR to the model or not.
- Random effects models assume $\delta_i \sim N(d, \tau^2)$, where τ^2 measures the heterogeneity of LOR across studies.

Bayesian approach: logistic regression

$$\begin{aligned} \text{FE model: } logit(p_{ik}) &= \mu_i + d * I(k = 2) \\ \text{RE model: } logit(p_{ik}) &= \mu_i + \delta_i * I(k = 2) \end{aligned}$$

- We adopt various prior distributions

Model name	Prior distribution
FE-vague	$\mu_i \sim N(0, 10^2)$ $d \sim N(0, 10^2)$
FE-informative	$\mu_i \sim N(0, 10^2)$ $d \sim N(d^*, 1)$
RE-vague	$\mu_i \sim N(0, 10^2)$ $\delta_i \sim N(d, \tau^2)$, $d \sim N(0, 10^2)$, $\tau \sim Unif(0, 2)$
RE-informative	$\mu_i \sim N(0, 10^2)$ $\delta_i \sim N(d, \tau^2)$, $d \sim N(d^*, 1)$, $\tau \sim Unif(0, 2)$
RE-shrinkage	$\mu_i \sim N(m, \tau_\mu^2)$, $m \sim N(0, 10^2)$, $\tau_\mu \sim Unif(0, 2)$ $\delta_i \sim N(d, \tau^2)$, $d \sim N(0, 10^2)$, $\tau \sim Unif(0, 2)$

Why Bayesian in meta-analysis?

- A Bayesian thinks of parameters as random, and thus having distributions.
- That is, Bayesian 95% credible intervals are more straightforward than frequentist 95% confidence intervals.
- Bayesian models incorporate all parameter uncertainties.
- **A Bayesian approach provides flexibility, enabling us to account for correlation induced by multi-arm trials.**
  - Bayesian hierarchical models
  - Prior specifications
- **A Bayesian approach offers probability-based interpretations, so we can easily calculate the ranking probabilities.**
  - Probability of being best, second best... or worst
  - Pr(effect size $<$ value $\mid$ Data)

HCQ confirmed case data

HCQ confirmed case data

study	ea	ec	na	nc
DukeHERO	3	6	683	676
Upenn	4	4	64	61
UofM	11	6	989	494
HenryFord	2	2	387	191
Mexico	1	6	62	65
India	11	12	211	203
EPICOS	21	23	224	211
Peru	5	3	34	31
Spain	1	1	137	116
Syed et al.	42	7	154	46

Stan Code for fixed effect model

```
ctelogit.stan = "
data{
  int NS;  // number of studies
  int ec[NS];  // number of events in control
  int ea[NS];  // number of events in active
  int nc[NS];  // sample size in control
  int na[NS];  // sample size in active
}
parameters{
  real lor;
  vector[NS] mu;
}
transformed parameters{
  vector[NS] logit_pc;
  vector[NS] logit_pa;
  vector[NS] pc;
  vector[NS] pa;
  //
  for(n in 1:NS) {
    logit_pc[n] = mu[n];
    logit_pa[n] = mu[n] + lor;
    //
    pc[n] = inv_logit(logit_pc[n]);
    pa[n] = inv_logit(logit_pa[n]);
  }
}
model{
  ec ~ binomial(nc, pc);
  ea ~ binomial(na, pa);
  //
  lor ~ normal(0, 100);
  mu ~ normal(0, 100);
}
```

```
generated quantities{
  vector[2*NS] log_lik;
  vector[NS] rdi;  // study-specific risk difference
  real rd;
  real phat_c;
  real phat_a;
  real abs_risk_reduction;
  real OR;
  real OR_re_less1;
  real OR_re_less05;
  real OR_re_gr2;
  int post_prob_less1;
  int post_prob_less05;
  int post_prob_gr2;

  for(n in 1:NS) {
    log_lik[n] = binomial_lpmf(ec[n] | nc[n], pc[n]);
    log_lik[NS+n] = binomial_lpmf(ea[n] | na[n], pa[n]);
  }
  for(n in 1:NS) {
    rdi[n] = pa[n]-pc[n];
  }
  rd = mean(rdi);
  phat_c = exp(mean(mu))/(1+exp(mean(mu)));
  phat_a = exp(mean(mu)+lor)/(1+exp(mean(mu)+lor));
  abs_risk_reduction = phat_a-phat_c;
  OR = exp(lor);

  OR_re_less1 = -OR + 1;
  OR_re_less05 = -OR + 0.5;
  OR_re_gr2 = OR - 2;

  post_prob_less1 = int_step(OR_re_less1);
  post_prob_less05 = int_step(OR_re_less05);
  post_prob_gr2 = int_step(OR_re_gr2);
}"
```

Stan fects model

```
htelogit.stan = "
data{
  int NS;  // number of studies
  int ec[NS];  // number of events in control
  int ea[NS];  // number of events in active
  int nc[NS];  // sample size in control
  int na[NS];  // sample size in active
}
parameters{
  real lor;
  real mu[NS];
  real<lower=0> tau;  // sd of delta
  real theta_tilde[NS];
}
transformed parameters{
  vector[NS] logit_pc;
  vector[NS] logit_pa;
  vector[NS] delta;
  vector[NS] pc;
  vector[NS] pa;
  //
  for(n in 1:NS) {
    delta[n] = lor + tau*tau*theta_tilde[n];
    logit_pc[n] = mu[n];
    logit_pa[n] = mu[n] + delta[n];
    //
    pc[n] = inv_logit(logit_pc[n]);
    pa[n] = inv_logit(logit_pa[n]);
  }
}
model{
  ec ~ binomial(nc, pc);
  ea ~ binomial(na, pa);
  //
  lor ~ normal(0, 100);
  mu ~ normal(0, 100);
  theta_tilde ~ normal(0, 1);
  tau ~ uniform(0,2);
}
```

```
generated quantities{
  vector[NS*2] log_lik;
  vector[NS] rdi;  // study-specific risk difference
  real rd;
  real phat_c;
  real phat_a;
  real abs_risk_reduction;
  real OR;
  real OR_re_less1;
  real OR_re_less05;
  real OR_re_gr2;
  int post_prob_less1;
  int post_prob_less05;
  int post_prob_gr2;

  for(n in 1:NS) {
    log_lik[n] = binomial_lpmf(ec[n] | nc[n], pc[n]);
    log_lik[NS+n] = binomial_lpmf(ea[n] | na[n], pa[n]);
  }
  for(n in 1:NS) {
    rdi[n] = pa[n]-pc[n];
  }
  rd = mean(rdi);
  OR = exp(lor);
  phat_c = exp(mean(mu))/(1+exp(mean(mu)));
  phat_a = exp(mean(mu)+lor)/(1+exp(mean(mu)+lor));
  abs_risk_reduction = phat_a-phat_c;

  OR_re_less1 = -OR + 1;
  OR_re_less05 = -OR + 0.5;
  OR_re_gr2 = OR - 2;

  post_prob_less1 = int_step(OR_re_less1);
  post_prob_less05 = int_step(OR_re_less05);
  post_prob_gr2 = int_step(OR_re_gr2);
}"
```

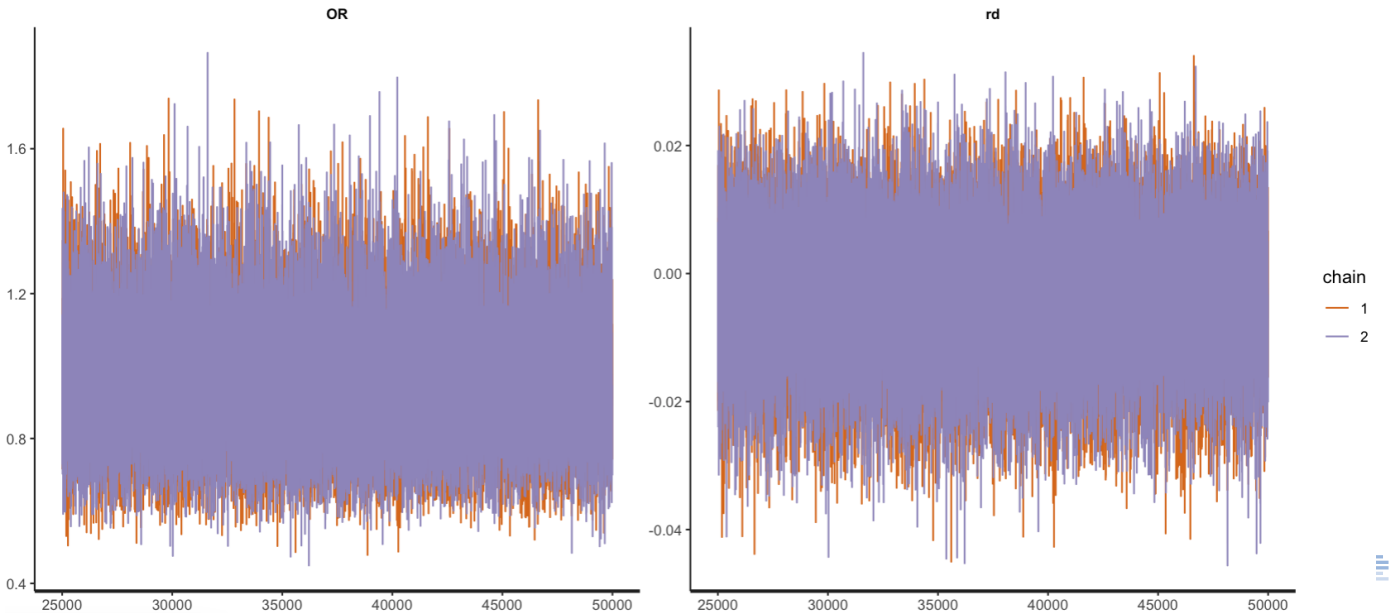
R code to run stan models

```
#######################
#### Run CTE_Logit ####
#######################

nchains <- 2
nthin <- 1
NS <- nrow(data)
ec <- data$ec
ea <- data$ea
nc <- data$nc
na <- data$na
niter <- 50000
nburnin <- 25000

stan.data <- list("NS" = NS, "ec" = ec, "ea" = ea, "nc" = nc, "na" = na)
stan.params <- c("OR", "rd", "phat_c", "phat_a", "abs_risk_reduction", "post_prob_less1", "post_prob_less05"
cte.fit <- NULL
cte.fit <- stan(model_code = ctelogit.stan, data = stan.data, pars = stan.params, seed = 752346,
                chains = nchains, iter = niter, warmup = nburnin, control = list(adapt_delta = 0.99))
print(cte.fit)$summary

## Check convergence
traceplot(cte.fit, pars=c("OR", "rd"))

#library(bayesplot)
cte.post.array <- as.array(cte.fit)
mcmc_trace(cte.post.array, pars = c("OR", "rd"))
```

R outputs: Fixed effect model

```
2 chains, each with iter=50000; warmup=25000; thin=1;
post-warmup draws per chain=25000, total post-warmup draws=50000.
```

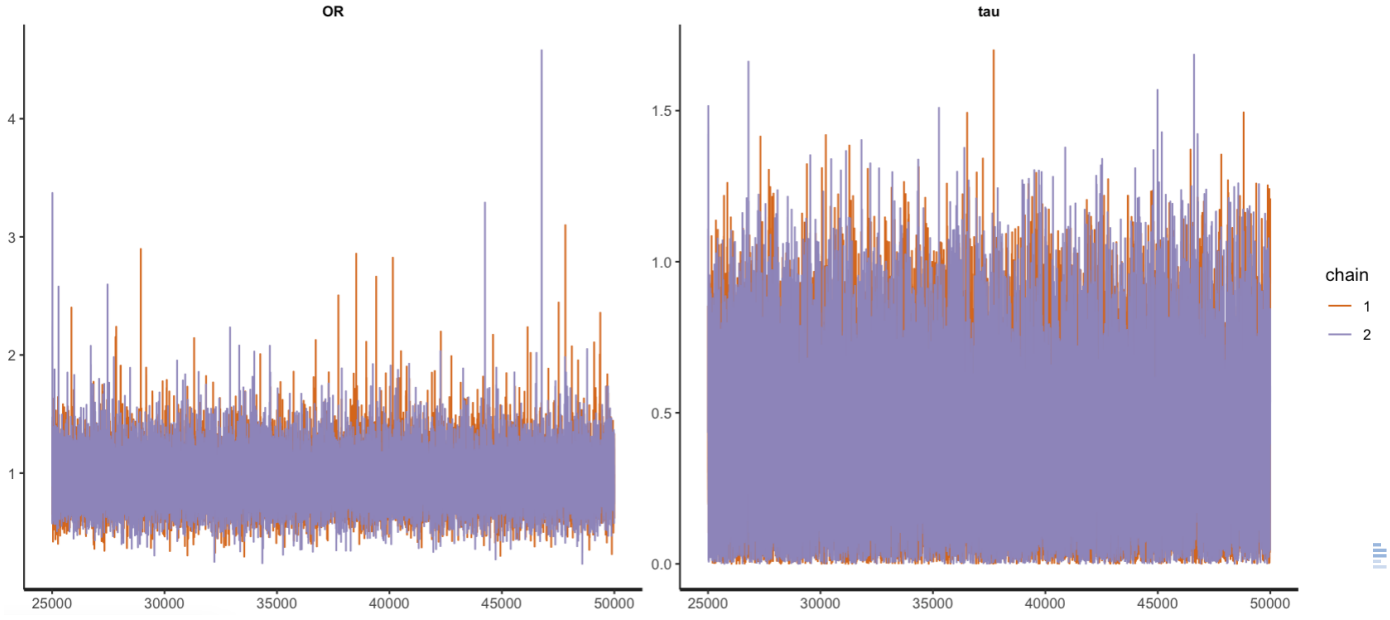
	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
OR	0.95	0.00	0.16	0.67	0.83	0.93	1.05	1.30	32552	1
rd	0.00	0.00	0.01	-0.02	-0.01	0.00	0.00	0.02	32216	1
phat_c	0.04	0.00	0.01	0.03	0.03	0.03	0.04	0.05	31608	1
phat_a	0.03	0.00	0.00	0.02	0.03	0.03	0.04	0.04	57437	1
abs_risk_reduction	0.00	0.00	0.01	-0.01	-0.01	0.00	0.00	0.01	31964	1
post_prob_less1	0.66	0.00	0.48	0.00	0.00	1.00	1.00	1.00	35316	1
post_prob_less05	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	40011	1
post_prob_gr2	0.00	NaN	0.00	0.00	0.00	0.00	0.00	0.00	NaN	NaN
lp__	-611.84	0.02	2.40	-617.49	-613.22	-611.49	-610.09	-608.17	20985	1



R outputs: Random effects model

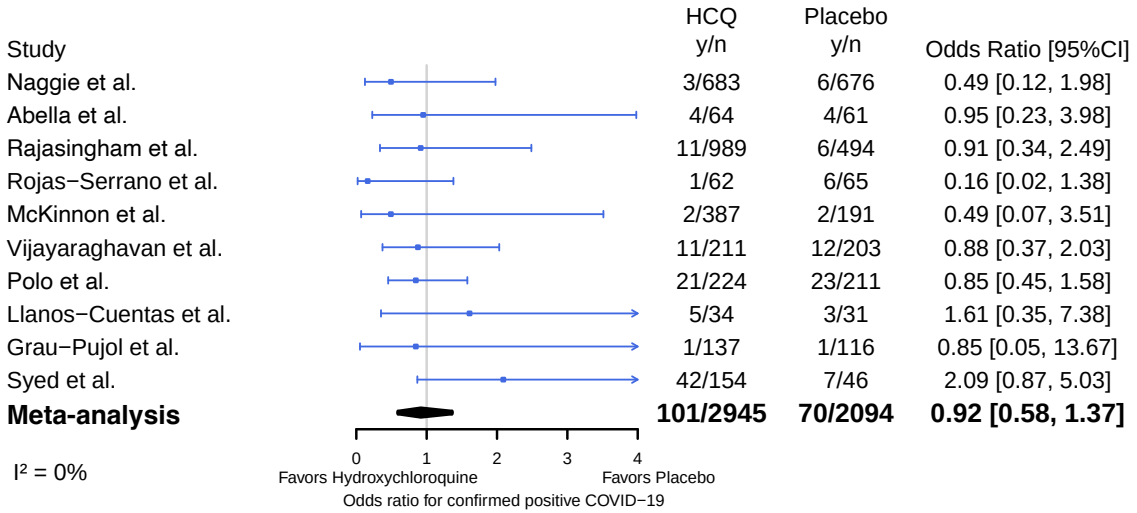
2 chains, each with iter=50000; warmup=25000; thin=1;
post-warmup draws per chain=25000, total post-warmup draws=50000.

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
tau	0.41	0.00	0.26	0.02	0.20	0.39	0.60	0.95	10494	1
OR	0.93	0.00	0.20	0.58	0.80	0.92	1.05	1.37	30567	1
rd	0.00	0.00	0.01	-0.02	-0.01	0.00	0.00	0.02	37798	1
phat_c	0.04	0.00	0.01	0.03	0.03	0.03	0.04	0.05	42388	1
phat_a	0.03	0.00	0.01	0.02	0.03	0.03	0.04	0.04	41107	1
abs_risk_reduction	0.00	0.00	0.01	-0.02	-0.01	0.00	0.00	0.01	31153	1
post_prob_less1	0.67	0.00	0.47	0.00	0.00	1.00	1.00	1.00	38570	1
post_prob_less05	0.01	0.00	0.09	0.00	0.00	0.00	0.00	0.00	23344	1
post_prob_gr2	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	25942	1
lp__	-617.33	0.03	3.71	-625.38	-619.65	-617.07	-614.72	-610.77	13976	1

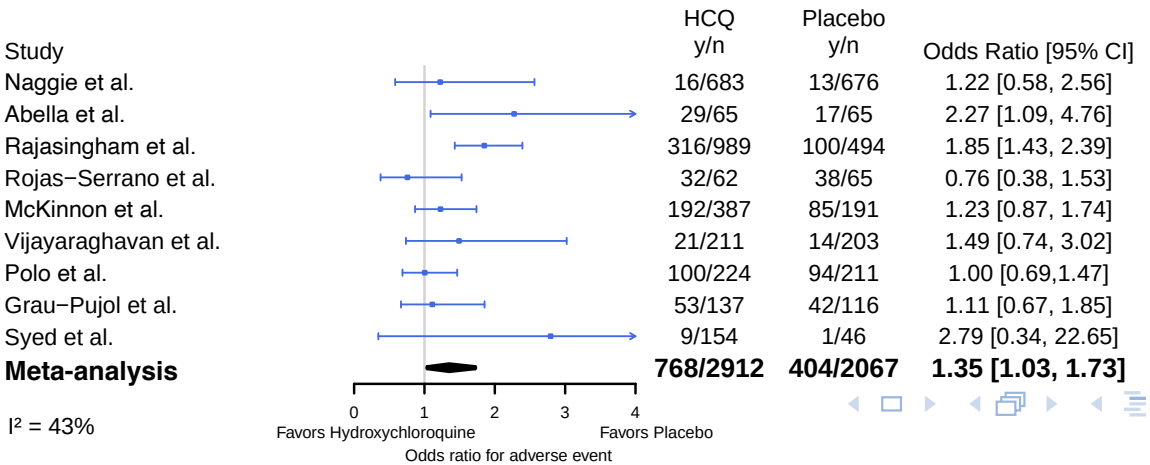


Forest plot

Confirmed positive COVID-19; $\hat{\tau} = 0.39$



Adverse event; $\hat{\tau} = 0.45$



Additional results

The Bayesian posterior probabilities of the OR less than 1 for the confirmed SARS-CoV-2 infection outcome (ie, the probability of HCQ favouring over placebo) was 0.67, while the posterior probability of OR less than 0.5 (ie, the probability that the odds of having a confirmed SARS-CoV-2 infection outcome in HCQ is less than a half of the odds in placebo) was 0.009. The posterior probability of the OR greater than 2 for the adverse event outcome (ie, the probability that the odds of having an adverse event in HCQ is greater than twice of the odds in placebo) was 0.004.

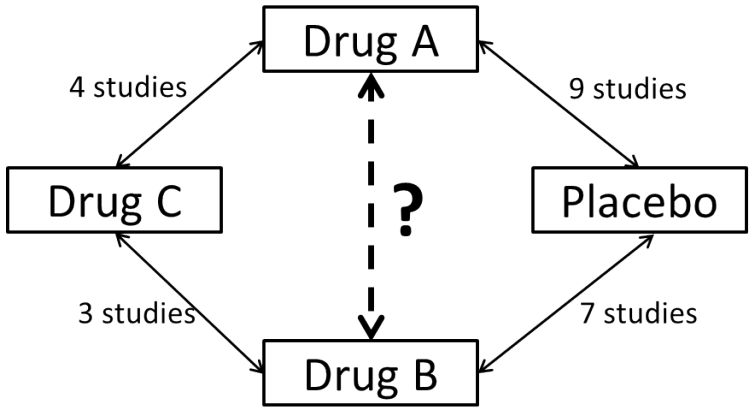
Exercise

Use the adverse event outcome data and run Bayesian meta-analysis models!

Outline

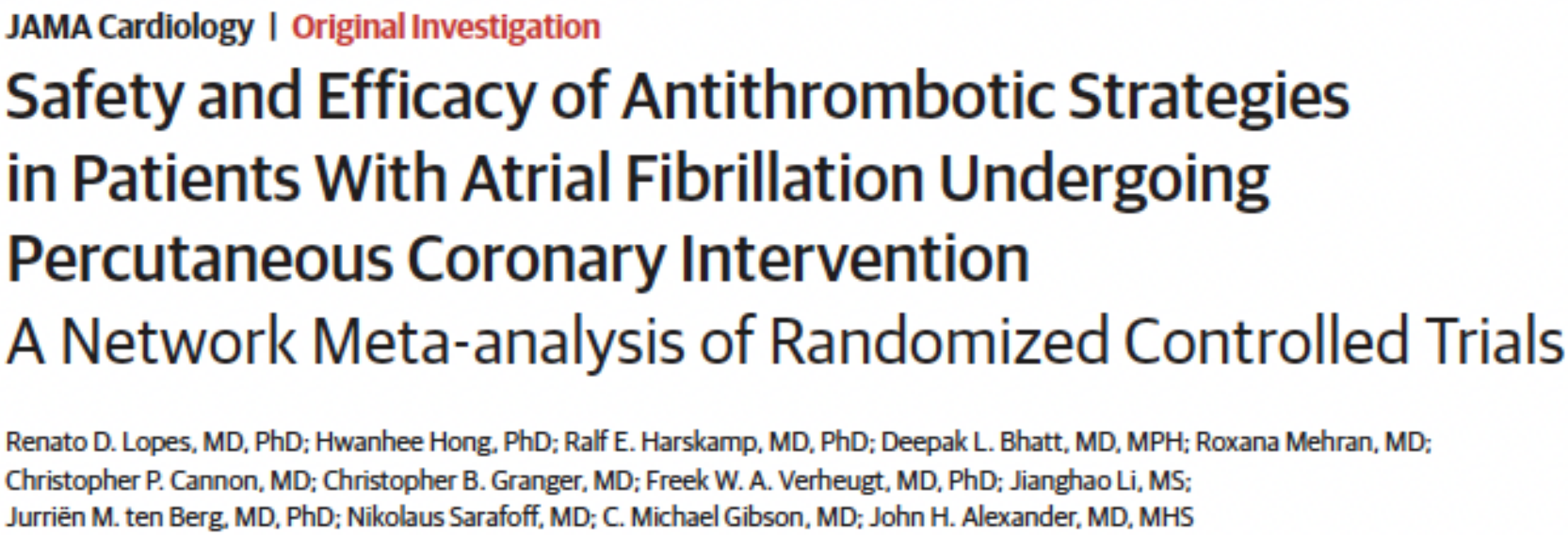
Network meta-analysis

- **Network meta-analysis (NMA)** is an extension of the pairwise meta-analysis to address the comparative effectiveness and safety of multiple (i.e., more than 2) interventions by combining all sources of data.



- NMA may or may not include head-to-head RCTs of the treatments of primary interest. In such cases, we must rely on indirect comparisons that use statistical techniques to incorporate the findings from multiple studies.
- Bayesian random effects NMA models are fitted.
  - "Random effects" are used to account for heterogeneity across studies.
  - "Bayesian" is a natural choice to model hierarchical evidence and correlation.
  - "Bayesian" gives us useful probability-based metrics to rank treatments.

Examples



JAMA Cardiology | Original Investigation

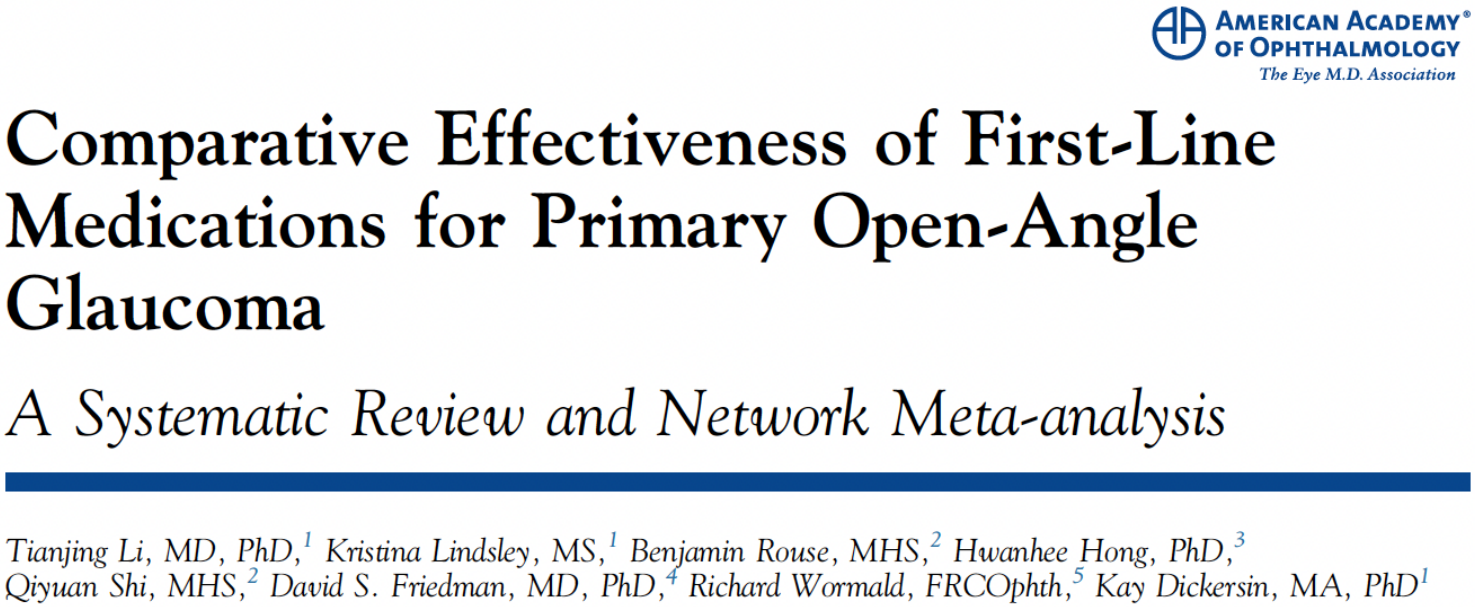
Safety and Efficacy of Antithrombotic Strategies in Patients With Atrial Fibrillation Undergoing Percutaneous Coronary Intervention

A Network Meta-analysis of Randomized Controlled Trials

Renato D. Lopes, MD, PhD; Hwanhee Hong, PhD; Ralf E. Harskamp, MD, PhD; Deepak L. Bhatt, MD, MPH; Roxana Mehran, MD; Christopher P. Cannon, MD; Christopher B. Granger, MD; Freek W. A. Verheugt, MD, PhD; Jianghao Li, MS; Jurriën M. ten Berg, MD, PhD; Nikolaus Sarafoff, MD; C. Michael Gibson, MD; John H. Alexander, MD, MHS

Lopes et al. (2019). *JAMA Cardiology* 4;747-755

Examples



AMERICAN ACADEMY OF OPHTHALMOLOGY
The Eye M.D. Association

Comparative Effectiveness of First-Line Medications for Primary Open-Angle Glaucoma

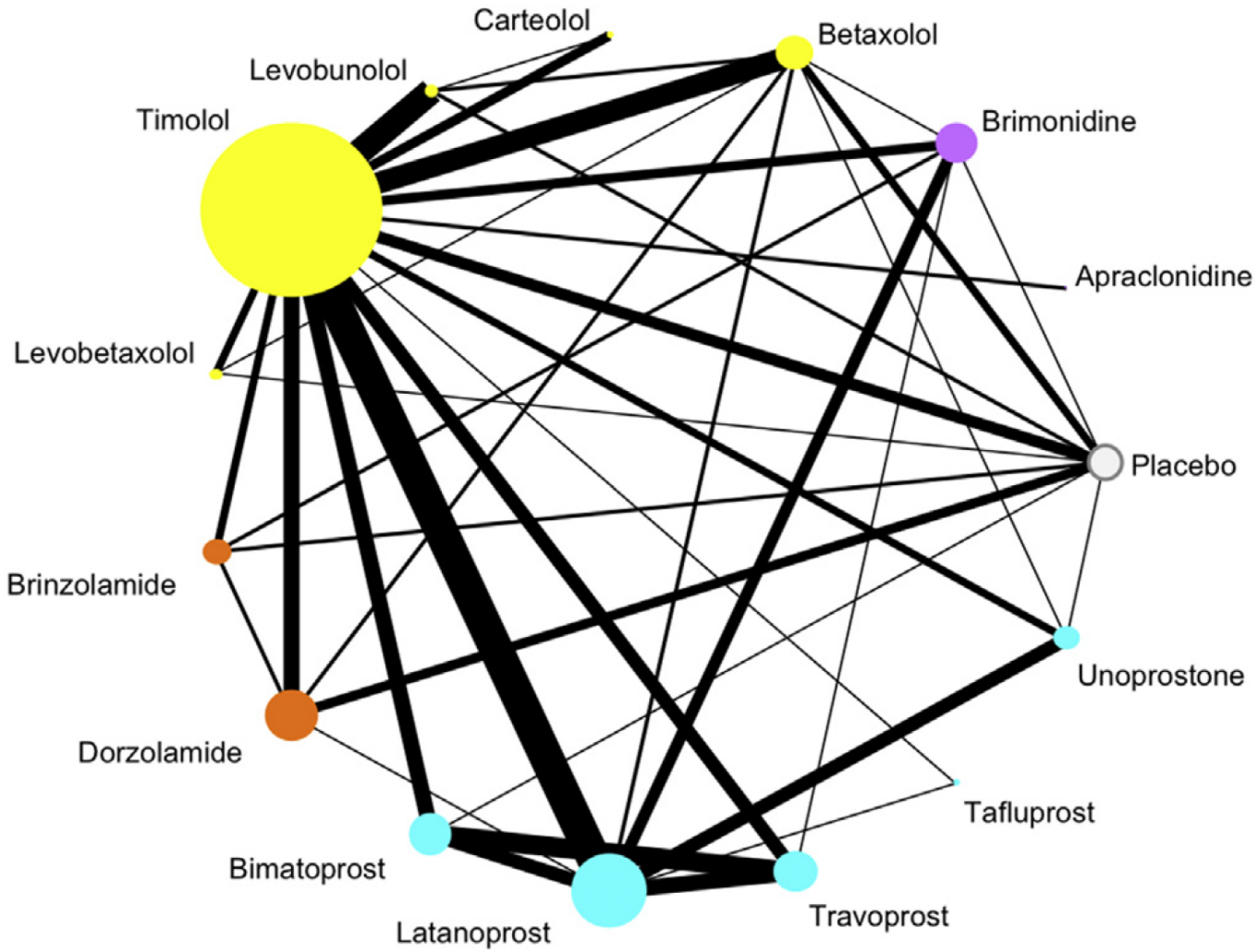
A Systematic Review and Network Meta-analysis

Tianjing Li, MD, PhD,[1] Kristina Lindsley, MS,[1] Benjamin Rouse, MHS,[2] Hwanhee Hong, PhD,[3] Qiyuan Shi, MHS,[2] David S. Friedman, MD, PhD,[4] Richard Wormald, FRCOphth,[5] Kay Dickersin, MA, PhD[1]

Li et al. (2015). *Ophthalmology* 123:129-140

NMA: POAG example

POAG: Primary Open-Angle Glaucoma



Li et al. (2015). *Ophthalmology* 123:129-140

NMA: POAG results

Table 2. Summary Estimates for Intraocular Pressure at 3 Months Derived from Network Meta-analysis of 114 Trials*

Placebo	−2.52 (−4.11; −0.94)	−3.59 (−4.29; −2.89)	−2.24 (−2.88; −1.59)	−3.44 (−4.46; −2.42)	−4.51 (−5.24; −3.8)	−3.7 (−4.24; −3.16)	−2.56 (−3.62; −1.52)	−2.42 (−3.23; −1.62)	−2.49 (−3.13; −1.85)	−5.61 (−6.29; −4.94)	−4.85 (−5.46; −4.24)	−4.83 (−5.54; −4.12)	−4.37 (−5.83; −2.94)	−1.91 (−2.67; −1.15)
2.52 (0.94; 4.11)	Apraclonidine	−1.08 (−2.65; 0.52)	0.28 (−1.29; 1.87)	−0.92 (−2.64; 0.81)	−2 (−3.58; −0.39)	−1.18 (−2.67; 0.31)	−0.05 (−1.81; 1.72)	0.09 (−1.55; 1.75)	0.03 (−1.54; 1.62)	−3.1 (−4.65; −1.53)	−2.33 (−3.85; −0.79)	−2.31 (−3.88; −0.73)	−1.85 (−3.85; 0.15)	0.61 (−0.99; 2.23)
3.59 (2.89; 4.29)	1.08 (−0.52; 2.65)	Brimonidine	1.35 (0.67; 2.04)	0.15 (−0.87; 1.17)	−0.92 (−1.69; −0.16)	−0.11 (−0.64; 0.42)	1.03 (−0.05; 2.1)	1.17 (0.4; 1.92)	1.1 (0.4; 1.8)	−2.02 (−2.69; −1.35)	−1.25 (−1.8; −0.72)	−1.24 (−1.92; −0.55)	−0.78 (−2.23; 0.65)	1.69 (0.92; 2.45)
2.24 (1.59; 2.88)	−0.28 (−1.87; 1.29)	−1.35 (−2.04; −0.67)	Betaxolol	−1.2 (−2.21; −0.19)	−2.28 (−3.01; −1.55)	−1.46 (−1.99; −0.94)	−0.32 (−1.38; 0.72)	−0.18 (−1.03; 0.65)	−0.25 (−0.91; 0.4)	−3.37 (−4.06; −2.7)	−2.61 (−3.2; −2.02)	−2.59 (−3.3; −1.89)	−2.13 (−3.58; −0.7)	0.33 (−0.42; 1.08)
3.44 (2.42; 4.46)	0.92 (−0.81; 2.64)	−0.15 (−1.17; 0.87)	1.2 (0.19; 2.21)	Carteolol	−1.08 (−2.07; −0.09)	−0.26 (−1.14; 0.61)	0.88 (−0.42; 2.16)	1.02 (−0.11; 2.14)	0.95 (−0.07; 1.96)	−2.17 (−3.16; −1.18)	−1.41 (−2.35; −0.47)	−1.39 (−2.4; −0.38)	−0.93 (−2.54; 0.66)	1.53 (0.47; 2.59)
4.51 (3.8; 5.24)	2 (0.39; 3.58)	0.92 (0.16; 1.69)	2.28 (1.55; 3.01)	1.08 (0.09; 2.07)	Levobunolol	0.81 (0.25; 1.38)	1.95 (0.86; 3.04)	2.09 (1.2; 2.99)	2.03 (1.28; 2.77)	−1.1 (−1.83; −0.36)	−0.33 (−0.99; 0.33)	−0.31 (−1.06; 0.44)	0.14 (−1.32; 1.6)	2.61 (1.79; 3.43)
3.7 (3.16; 4.24)	1.18 (−0.31; 2.67)	0.11 (−0.42; 0.64)	1.46 (0.94; 1.99)	0.26 (−0.61; 1.14)	−0.81 (−1.38; −0.25)	Timolol	1.14 (0.19; 2.07)	1.28 (0.56; 1.99)	1.21 (0.69; 1.73)	−1.91 (−2.38; −1.44)	−1.15 (−1.5; −0.79)	−1.13 (−1.63; −0.63)	−0.67 (−2.02; 0.67)	1.79 (1.18; 2.41)
2.56 (1.52; 3.62)	0.05 (−1.72; 1.81)	−1.03 (−2.1; 0.05)	0.32 (−0.72; 1.38)	−0.88 (−2.16; 0.42)	−1.95 (−3.04; −0.86)	−1.14 (−2.07; −0.19)	Levobetaxolol	0.14 (−1.03; 1.32)	0.07 (−0.99; 1.14)	−3.05 (−4.09; −1.99)	−2.28 (−3.28; −1.27)	−2.27 (−3.32; −1.19)	−1.81 (−3.45; −0.18)	0.66 (−0.45; 1.78)
2.42 (1.62; 3.23)	−0.09 (−1.75; 1.55)	−1.17 (−1.92; −0.4)	0.18 (−0.65; 1.03)	−1.02 (−2.14; 0.11)	−2.09 (−2.99; −1.2)	−1.28 (−1.99; −0.56)	−0.14 (−1.32; 1.03)	Brinzolamide	−0.07 (−0.85; 0.71)	−3.19 (−4.03; −2.36)	−2.42 (−3.18; −1.66)	−2.41 (−3.26; −1.55)	−1.95 (−3.47; −0.44)	0.52 (−0.39; 1.43)
2.49 (1.85; 3.13)	−0.03 (−1.62; 1.54)	−1.1 (−1.8; −0.4)	0.25 (−0.4; 0.91)	−0.95 (−1.96; 0.07)	−2.03 (−2.77; −1.28)	−1.21 (−1.73; −0.69)	−0.07 (−1.14; 0.99)	0.07 (−0.71; 0.85)	Dorzolamide	−3.12 (−3.8; −2.44)	−2.36 (−2.95; −1.76)	−2.34 (−3.05; −1.63)	−1.88 (−3.32; −0.45)	0.58 (−0.19; 1.36)
5.61 (4.94; 6.29)	3.1 (1.53; 4.65)	2.02 (1.35; 2.69)	3.37 (2.7; 4.06)	2.17 (1.18; 3.16)	1.1 (0.36; 1.83)	1.91 (1.44; 2.38)	3.05 (1.99; 4.09)	3.19 (2.36; 4.03)	3.12 (2.44; 3.8)	Bimatoprost	0.77 (0.27; 1.26)	0.78 (0.26; 1.3)	1.24 (−0.18; 2.65)	3.71 (2.97; 4.44)
4.85 (4.24; 5.46)	2.33 (0.79; 3.85)	1.25 (0.72; 1.8)	2.61 (2.02; 3.2)	1.41 (0.47; 2.35)	0.33 (−0.33; 0.99)	1.15 (0.79; 1.5)	2.28 (1.27; 3.28)	2.42 (1.66; 3.18)	2.36 (1.76; 2.95)	−0.77 (−1.26; −0.27)	Latanoprost	0.02 (−0.5; 0.53)	0.48 (−0.91; 1.83)	2.94 (2.33; 3.55)
4.83 (4.12; 5.54)	2.31 (0.73; 3.88)	1.24 (0.55; 1.92)	2.59 (1.89; 3.3)	1.39 (0.38; 2.4)	0.31 (−0.44; 1.06)	1.13 (0.63; 1.63)	2.27 (1.19; 3.32)	2.41 (1.55; 3.26)	2.34 (1.63; 3.05)	−0.78 (−1.3; −0.26)	−0.02 (−0.53; 0.5)	Travoprost	0.46 (−0.98; 1.87)	2.92 (2.17; 3.68)
4.37 (2.94; 5.83)	1.85 (−0.15; 3.85)	0.78 (−0.65; 2.23)	2.13 (0.7; 3.58)	0.93 (−0.66; 2.54)	−0.14 (−1.6; 1.32)	0.67 (−0.67; 2.02)	1.81 (0.18; 3.45)	1.95 (0.44; 3.47)	1.88 (0.45; 3.32)	−1.24 (−2.65; 0.18)	−0.48 (−1.83; 0.91)	−0.46 (−1.87; 0.98)	Tafluprost	2.46 (1.01; 3.95)
1.91 (1.15; 2.67)	−0.61 (−2.23; 0.99)	−1.69 (−2.45; −0.92)	−0.33 (−1.08; 0.42)	−1.53 (−2.59; −0.47)	−2.61 (−3.43; −1.79)	−1.79 (−2.41; −1.18)	−0.66 (−1.78; 0.45)	−0.52 (−1.43; 0.39)	−0.58 (−1.36; 0.19)	−3.71 (−4.44; −2.97)	−2.94 (−3.55; −2.33)	−2.92 (−3.68; −2.17)	−2.46 (−3.95; −1.01)	Unoprostone

*Posterior means (95% Bayesian credible intervals) are calculated by column − row under the Lu and Ades[22] homogeneous random effects model assuming consistency. Mean difference <0 favors the drug in the column, and mean difference >0 favors the drug in the row.
Color coding:

White	Placebo/vehicle/no treatment
Purple	Alpha-2 adrenergic agonist
Yellow	Beta-blocker
Orange	Carbonic anhydrase inhibitor
Turquoise	Prostaglandin analog

Li et al. (2015). *Ophthalmology* 123:129-140

NMA: POAG results

Ranks	Placebo	Unoprostone	Betaxolol	Brinzolamide	Dorzolamide	Apraclonidine	Levobetaxolol	Carteolol	Brimonidine	Timolol	Tafluprost	Levobunolol	Travoprost	Latanoprost	Bimatoprost
1	0	0	0	0	0	0	0	0	0	0	0.042	0.002	0.001	0.001	0.953
2	0	0	0	0	0	0.001	0	0	0	0	0.155	0.082	0.356	0.361	0.046
3	0	0	0	0	0	0.001	0	0.001	0	0	0.081	0.129	0.361	0.426	0.001
4	0	0	0	0	0	0.003	0	0.008	0.003	0	0.167	0.424	0.215	0.18	0
5	0	0	0	0	0	0.014	0.002	0.069	0.067	0.074	0.333	0.343	0.065	0.032	0
6	0	0	0	0	0	0.031	0.006	0.178	0.248	0.437	0.079	0.018	0.002	0	0
7	0	0	0	0.001	0.001	0.033	0.014	0.17	0.339	0.379	0.061	0.002	0	0	0
8	0	0	0.002	0.019	0.013	0.082	0.063	0.379	0.285	0.102	0.055	0	0	0	0
9	0	0.006	0.033	0.143	0.152	0.206	0.245	0.136	0.055	0.006	0.018	0	0	0	0
10	0	0.021	0.101	0.214	0.283	0.125	0.21	0.037	0.004	0	0.005	0	0	0	0
11	0	0.049	0.194	0.219	0.281	0.095	0.147	0.013	0	0	0.002	0	0	0	0
12	0	0.108	0.289	0.191	0.178	0.097	0.13	0.006	0	0	0.001	0	0	0	0
13	0	0.265	0.285	0.144	0.075	0.119	0.111	0.002	0	0	0	0	0	0	0
14	0.001	0.551	0.097	0.068	0.017	0.192	0.072	0.001	0	0	0	0	0	0	0
15	0.999	0	0	0	0	0.001	0	0	0	0	0	0	0	0	0

Histogram of ranking probabilities for each drug (the warmer the color, the better the rank).

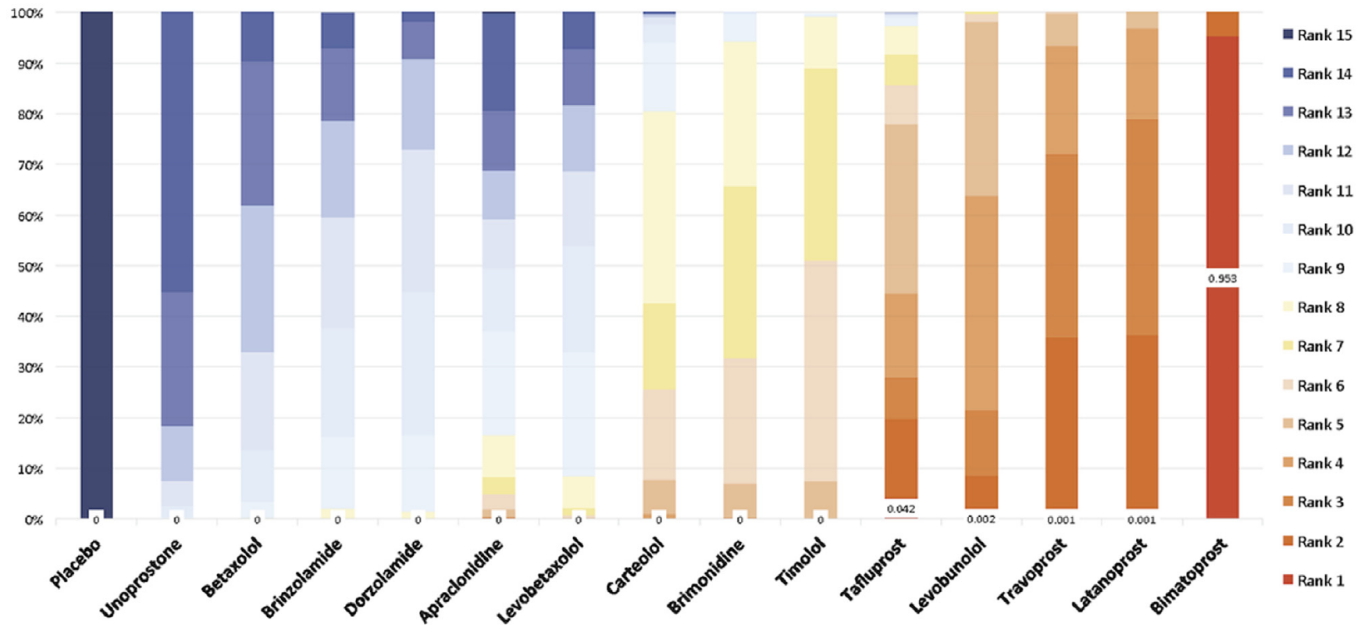


Figure 4. Ranking probabilities for any drug at any position.

Li et al. (2015). *Ophthalmology* 123:129-140

Summary

- MA and NMA are statistical methods to combine/integrate/synthesize multiple studies to generate relative effect estimates between two interventions.
- MA and NMA improve our understanding of comparative effectiveness of multiple interventions by leveraging multiple data sources.
- Statistical model specifications should be carefully reported.
- The effect estimates and associated 95% confidence/credible intervals should always be presented.
- The rankings based on probabilities are straightforward and attractive, BUT could be over-interpreted and misleading when making decisions.
- Many methods for meta-analysis are actively developing for specific topics such as generalizability, treatment effect heterogeneity, historical controls in RCTs.

Thank you!

 www.HwanheeHong.com