Borrowing strength across informative and sparse networks of interventions

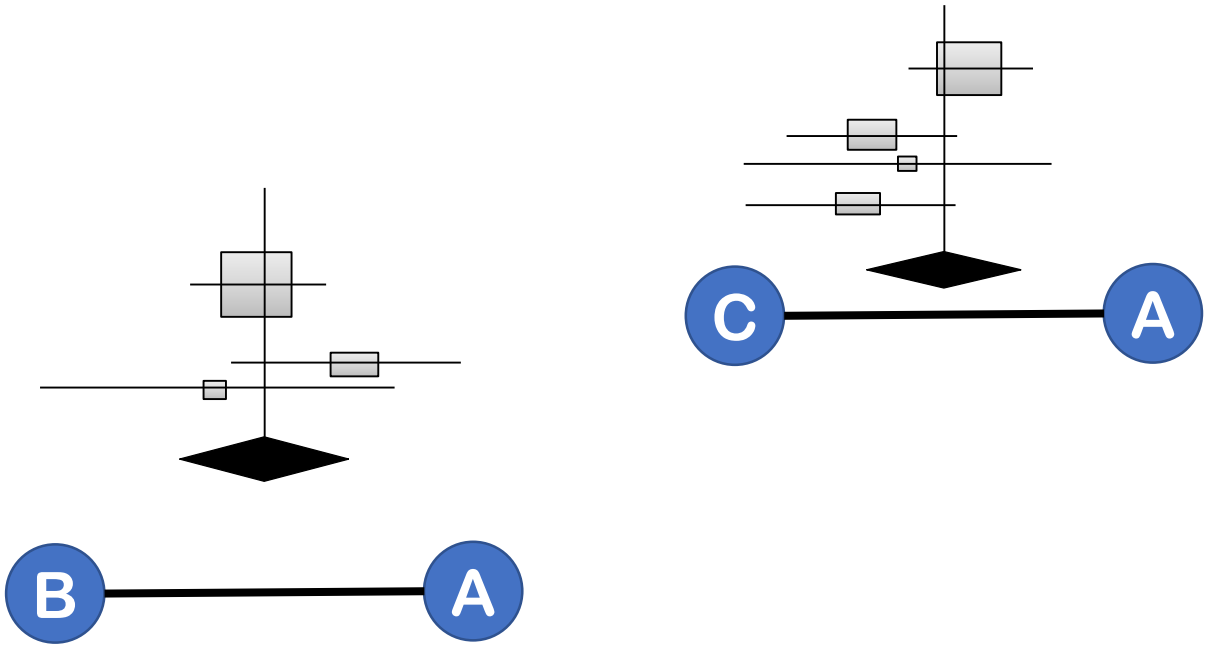
Theodoros Evrenoglou

Silvia Metelli, Johannes-Schreider Thomas, Spyridon Siafis, Stefan Leucht, Anna Chaimani

Université Paris Cité, Research Center of Epidemiology & Statistics (CRESS-UMR1153), Inserm, France

Work in progress seminar
MRC Clinical Trials Unit, University College London
21/06/2022

Introduction to Network Meta-Analysis (NMA)

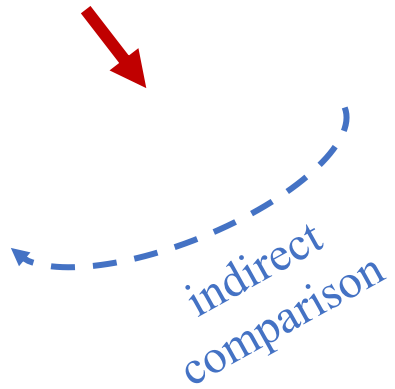


Introduction to Network Meta-Analysis (NMA)

Advantages of NMA

- ✓ Allows the comparison of treatments that have never been compared directly in individual studies.
- ✓ Combines both direct and indirect evidence resulting into estimates with highest precision.
- ✓ Allows for a relative ranking of the competing treatments.

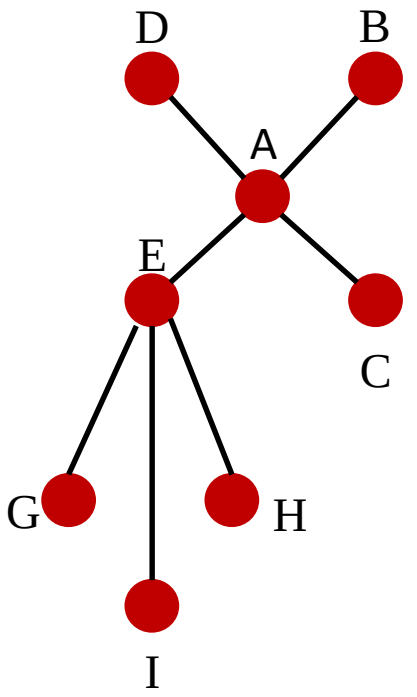
Why to perform a NMA?



Sparse data in NMA

- The issue refers to poorly connected networks with limited direct comparisons and few studies to inform them
- Such networks are arising for sensitive subgroups of the population (e.g. children, elder patients, individuals with multimorbidity)
- Statistical challenges in the case of sparse networks
  - large sample approximations fail
  - NMA estimates are expected to be imprecise and biased
  - the formal evaluation of NMA assumptions (transitivity, consistency) is challenging
- Overall, there is a lack of robustness and limited reliability to the NMA estimates when analyzing sparse treatment networks

An example of a sparse network



Frequency of sparse networks in the literature

- In a review of 186 NMAs it was found that
  - the median number of studies in a NMA is 21
  - the median number of studies per comparison is 2

- In a sample of 1236 networks of RCTs with at least 4 treatments it was found that 92 (7.4%) had more treatments than studies

- Sparse networks are frequent in the literature

Nikolakopoulou A et al. (2014) Characteristics of Networks of Interventions: A Description of a Database of 186 Published Networks. https://doi.org/10.1371/journal.pone.0086754

Yoon JH, Dias S, Hahn S. A method for assessing robustness of the results of a star-shaped network meta-analysis under the unidentifiable consistency assumption. BMC Med Res Methodol. 2021;21(1):113. doi:10.1186/s12874-021-01290-1

Aim of our work

- To propose a framework suitable for NMA of sparse networks in order to
  - improve the precision in the estimation
  - increase the reliability of the final NMA estimates

- Our idea is to use external evidence and to construct informative priors for the analysis of the sparse network

Motivating dataset

- A dataset with different antipsychotics and different populations of patients
- Across all populations, only the population of the general patients (GP) defines an informative network

- **<u>Clinical question:</u> Investigate the effectiveness of the available antipsychotics for the overall reduction of schizophrenia symptoms**

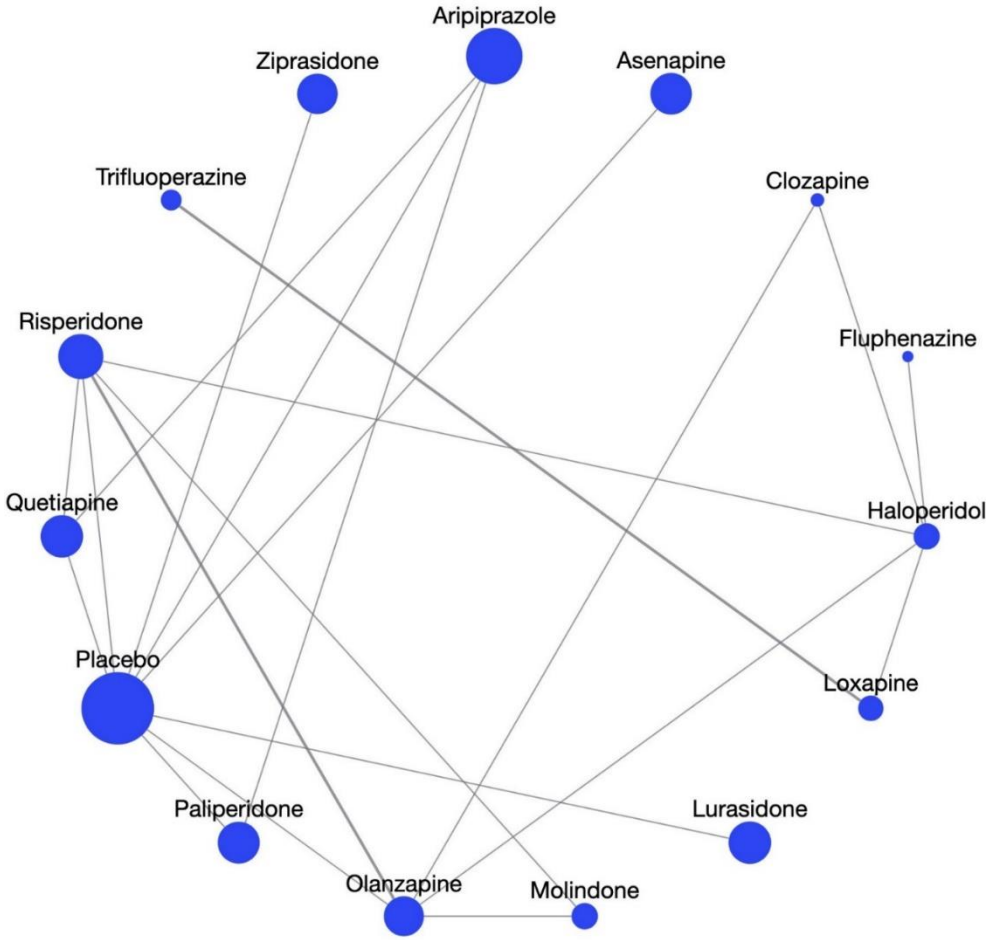
- We will illustrate our method using the informative network of GP and the sparse network of children and adolescents (CA)

Data description-Sparse network

- A sparse network of 14 antipsychotics and Placebo
  - **Population of interest**: Children and adolescents (CA)
  - **Outcome of interest**: Reduction in overall schizophrenia symptoms (continuous outcome)

- 21 direct comparisons 19 studies in the network
  - 90% of the direct comparisons are informed by 1 study

- **Year range: 1973-2017 (48 years)** → We cannot rely on new evidence



Objective:
Obtain reliable NMA estimates for the effectiveness of the antipsychotics in the population of CA

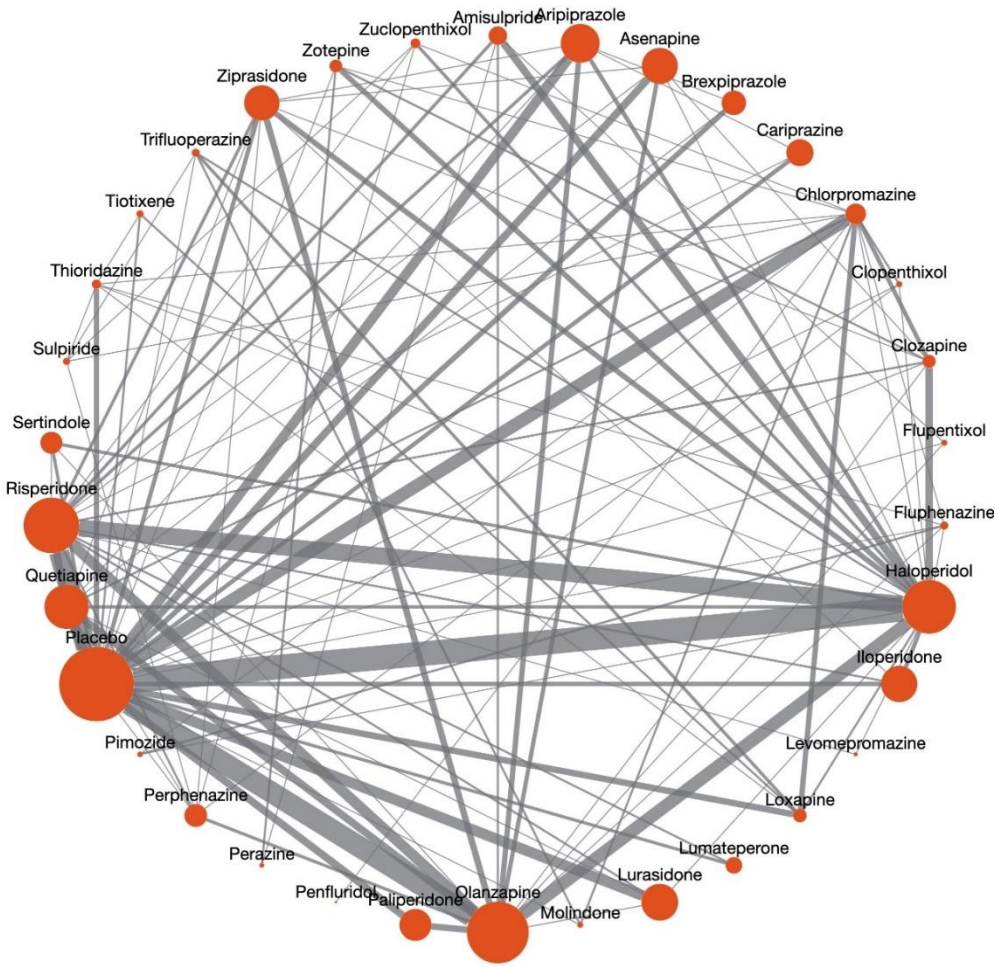
Krause M, Zhu Y, Huhn M, et al. Efficacy, acceptability, and tolerability of antipsychotics in children and adolescents with schizophrenia: A network meta-analysis. Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol. 2018;28(6):659-674. doi:10.1016/j.euroneuro.2018.03.008

The dense network of general patients

- An informative network of 33 antipsychotics and Placebo
  - **<u>Population of interest</u>**: General patients (GP)
  - **<u>Outcome of interest</u>**: Reduction in overall schizophrenia symptoms (continuous outcome)

- 116 direct comparisons and 255 studies in the network

- **Year range: 1967-2021 (54 years)**



<u>Main goal</u>

Borrow strength from the network of GP to analyse the network of CA

Huhn M, Nikolakopoulou A, Schneider-Thoma J, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. Lancet Lond Engl. 2019;394(10202):939-951. doi:10.1016/S0140-6736(19)31135-3

How to borrow strength?

Naïve synthesis:

1. Analyse the 2 populations together, or
2. Use **directly** NMA estimates of GP as prior information for CA

Assumptions

- We assume no population difference
- We assume that they are equivalent sources

implausible assumptions

A more plausible assumption:

The two populations share enough common for sharing of information to make sense, but they are not equivalent

We use a two stage approach:

✓ At the first stage we extrapolate the results of the dense network of GP to CA

✓ At the second stage we use the predictions of the extrapolations to form informative priors and we analyze the network of CA

Notation and general settings

- Let y_{it_k} denote the observed mean for the treatment k in study i

Within-studies assumption: $y_{it_k} \sim N(\theta_{it_k}, sd^2_{it_k})$

- Let $\delta_{i,t_1t_k} = \frac{\theta_{it_k} - \theta_{it_1}}{sd_i^{pooled}}$ denote the SMD between treatment k versus the baseline treatment 1 in study i

Across-studies assumption: $\delta_i \sim N_{K_i-1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

arms in study i

matrix structure: τ^2 in the diagonal, $\frac{\tau^2}{2}$ in the off diagonal

- Under the consistency assumption, $\mu_{t_kt_l} = \mu_{t_bt_l} - \mu_{t_bt_k}$

1st stage: Extrapolating GP results to the CA

- At this stage we only analyze the network of GP
  - we reduce the network to include only the 15 interventions that exists for CA
  - we use a modified NMA model to analyze GP

- We add a scale parameter $\boldsymbol{w}$ at the within-studies assumption. This parameter aims to inflate the variance of per study mean in GP

$$y_{it_k} \sim N\left(\theta_{it_k}, \frac{sd_{ik}^2}{w_i}\right), \mathrm{w_i} \in (0,1]$$

e.g. downweight the studies with high risk of bias

- We add a location parameter $\boldsymbol{\beta}$ that aims to shift the original distribution of the SMD's in GP towards the distribution of CA

$$\delta_i \sim N_{K_i-1}(\boldsymbol{\mu} - \boldsymbol{\beta}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\beta} = \left(\beta_{t_1,t_2}, \beta_{t_1,t_3}, \dots, \beta_{t_1 t_{Ki}}\right), \beta_{t_1 t_{K_i}} = \beta_{t_{K_i}} - \beta_{t_1}$$

2nd stage: NMA for CA using informative priors

- By fitting the modified NMA model we will obtain extrapolated SMD estimates $\hat{\mu}_{t_k t_l}^{\mathrm{GP}}, \forall t_k, t_l$
- The predictive distributions of these estimates are defined as $m_{t_k t_l}^{GP} \sim N\left(\hat{\mu}_{t_k t_l}^{GP}, \tau^{2,\mathrm{GP}}\right)$
- The predictive distributions are used as informative prior distributions for CA

$$\mu_{t_b t_k}^{CA} \sim N\left(\hat{m}_{t_k t_l}^{GP}, var\left(\hat{m}_{t_k t_l}^{GP}\right)\right)$$

- This is the key difference between our approach and the standard NMA model which places non-informative priors for μ's

Informing the location and the scale parameters

- The parameters $\boldsymbol{w}$ and $\boldsymbol{\beta}$ can be assumed either being fixed values or following a distribution
  - we assume the latter more conservative approach

- For the location parameter $\boldsymbol{\beta}$ we consider two approaches
  - a data based approach
  - prior elicitation from expert opinion

- For the scale parameter $\boldsymbol{w}$ the prior distribution is related to the amount of downweight that we want to apply according to a specific criterion (e.g. high risk of bias)
  - dividing within-study variances with w_i is a special case of the power prior method
  - usual choices of priors are the $Beta(\varepsilon_1, \varepsilon_2)$ or the $Unif(\varepsilon_1, \varepsilon_2)$

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015) The power prior: theory and applications. Statist. Med., 34: 3724– 3749. doi: 10.1002/sim.6728..

Data based approach for $\beta's$

- We assume that $u_{it_k} = \frac{\theta_{it_k}}{sd_i^{pooled}}$ for every study i and treatment t_k in both networks

- We obtain a treatment-specific average standardized mean through a meta-analysis in the studies for every network

$$u_{it_k} \sim N\left(\xi_{t_k}^{GP}, \sigma^{2,\mathrm{GP}}\right), i = 1,2,\dots,n_1$$

$$u_{it_k} \sim N\left(\xi_{t_k}^{CA}, \sigma^{2,\mathrm{CA}}\right), i = n_1+1, n_1+2, \dots, N$$

for simplicity we can assume that $\sigma^{2,GP} = \sigma^{2,\mathrm{CA}}$

- The difference $d_{t_k} = \xi_{t_k}^{GP} - \xi_{t_k}^{CA}$ is used as prior for β_{t_k}, $\beta_{t_k} \sim N(\hat{d}_{t_k}, var(\hat{d}_{t_k}))$

Elicitation of expert opinion for $\beta's$

- We prepared a questionnaire and circulated it to experts with experience in schizophrenia treatment

- We asked the experts to give us directly an estimate of $\xi_{t_k}^{\mathrm{CA}}$ given the $\xi_{t_k}^{GP}$, they were also asked to give us
  - an uncertainty measure around $\xi_{t_k}^{\mathrm{CA}}$ (i.e. standard deviation)
  - their confidence to the number that they did provide for $\xi_{t_k}^{\mathrm{CA}}$ and the uncertainty measure

- We received in total 22 responses which were averaged using a similar to the data-based approach meta-analysis model
  - we inflate the uncertainty measure of each expert using the confidence that they did provide
  - in this way expert opinions with higher confidence are having an increased weight

Implementation of the methodology

- We used both the data-based approach and the expert opinion to construct the priors for the location parameter $\boldsymbol{\beta}$

- In terms of the scale parameter $\boldsymbol{w}$ we apply 3 different downweighting schemes for the studies in GP
  - no downweight (No DW)
  - downweight for the studies with high risk of bias (RoB DW)
  - use of the full network for GP (34 treatments) and perform downweight to all studies that contain at least one treatment arm that is not common between the two networks (NCT DW)

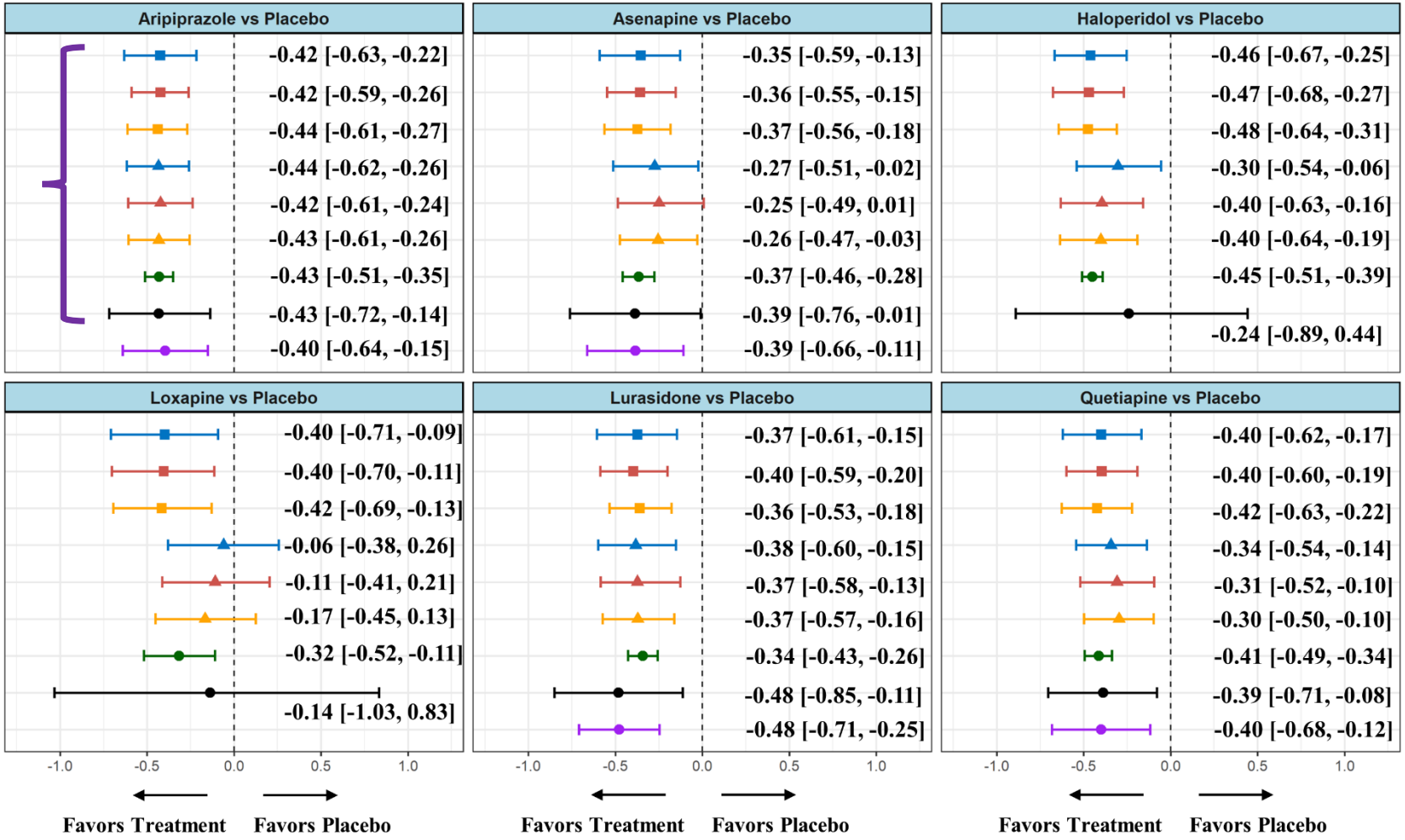
- For both downweighting schemes we assume that $w_i = w \sim Beta(3,3)$

Results for CA

- Data based $\boldsymbol{\beta}$ – No DW
- Data based $\boldsymbol{\beta}$ – RoB DW
- Data based $\boldsymbol{\beta}$ – NCT DW
- Expert opinion based $\boldsymbol{\beta}$ – No DW
- Expert opinion based $\boldsymbol{\beta}$ – RoB DW
- Expert opinion based $\boldsymbol{\beta}$ – NCT DW
- Naïve pooling
- NMA with non-informative priors
- Direct comparison

Prior for downweight parameter w: $w \sim Beta(3,3)$

No-DW: No downweight
RoB-DW: Risk of bias downweight
NCT- Non common treatment downweight

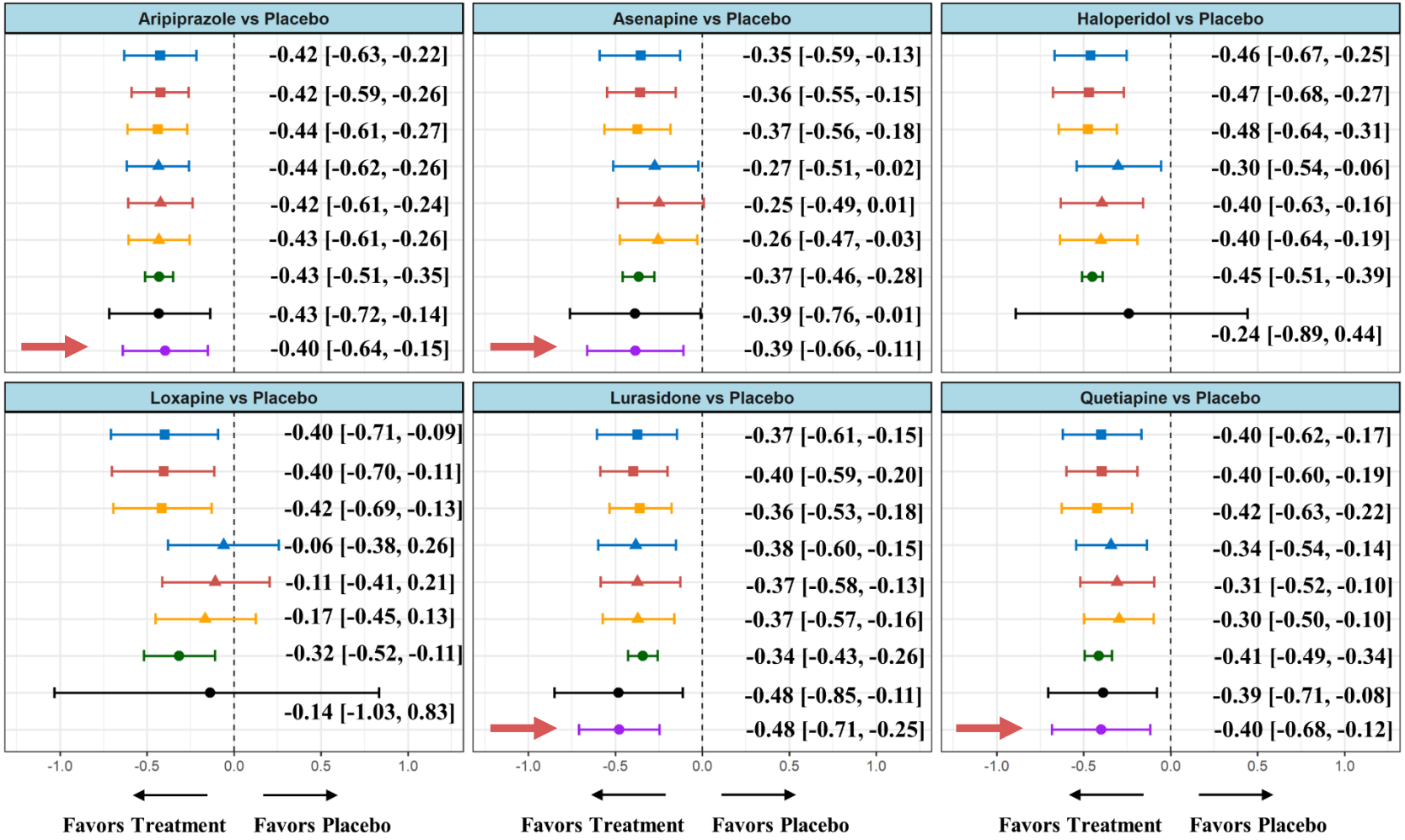


Results for CA

- Data based $\boldsymbol{\beta}$ – No DW
- Data based $\boldsymbol{\beta}$ – RoB DW
- Data based $\boldsymbol{\beta}$ – NCT DW
- Expert opinion based $\boldsymbol{\beta}$ – No DW
- Expert opinion based $\boldsymbol{\beta}$ – RoB DW
- Expert opinion based $\boldsymbol{\beta}$ – NCT DW
- Naïve pooling
- NMA with non-informative priors
- Direct comparison

Prior for downweight parameter w: $w \sim Beta(3,3)$

No-DW: No downweight
RoB-DW: Risk of bias downweight
NCT- Non common treatment downweight



Discussion

- In this work we proposed a framework for analyzing sparse networks by using informative priors
  - the precision and the reliability of the estimates is improved as they consider multiple source of evidence
  - our method can applied for sharing information between any dense network P_2 and a sparse network P_1
- There are limitations in our work
  - the heterogeneity estimation is still based on the data coming from the sparse network
  - for the data-based approach we need to use the data two times
  - the experts found it hard to imagine the differences between the GP and CA
- To conclude sharing of information seems to facilitate the estimation of treatment effects in sparse networks
  - Extensive sensitivity analysis across different choices of prior distributions should always take place to investigate the robustness of the results across different analysis schemes

THANK YOU!

QUESTIONS?