# Designing and dissecting the geometry of randomized evidence

SCT 2012

John P.A. Ioannidis, MD, DSc

C.F. Rehnborg Chair in Disease Prevention
Professor of Medicine and Professor of Health Research and Policy
Director, Stanford Prevention Research Center
Stanford University School of Medicine
Professor of Statistics (by courtesy)
Stanford University School of Humanities and Sciences

# Disclosures

- In my dreams I am the CEO of MMM (Make More Money, Inc.)
- My company has successfully developed a new drug that is probably a big loser, but I want to make big money
- At best, the new drug may be modestly effective for one or two diseases/indications for one among many outcomes (most of them irrelevant to patients)
- If I test my drug in a study, even for this one or two indications, it may seem not to be worth it
- But still, I want to make big money
- What should I do?

# The answer

- Run many studies with many outcomes on each of many different indications
- Ideally run trials against placebo (this is the gold standard for regulatory agencies) or straw man comparators, but registry studies or even electronic records would do, if need be
- Test 10 indications and 10 outcomes for each, just by chance you get 5 indications with statistically significant beneficial results
- A bit of selective outcome and analysis will help present "positive" results for 7-8, maybe even for all 10 indications
- There are systematic reviewers out there who will perform a systematic review based on the published data **SEPARATELY** for each indication proving the drug works for all 10 indications
- With $ 1 billion market share per approved indication, we can make $ 10 billion a year out of an (almost) totally useless drug

# We probably all agree

- It is stupid to depend on the evidence of a single study

- when there are many studies and a meta-analysis thereof on the same treatment comparison and same indication

# Similarly

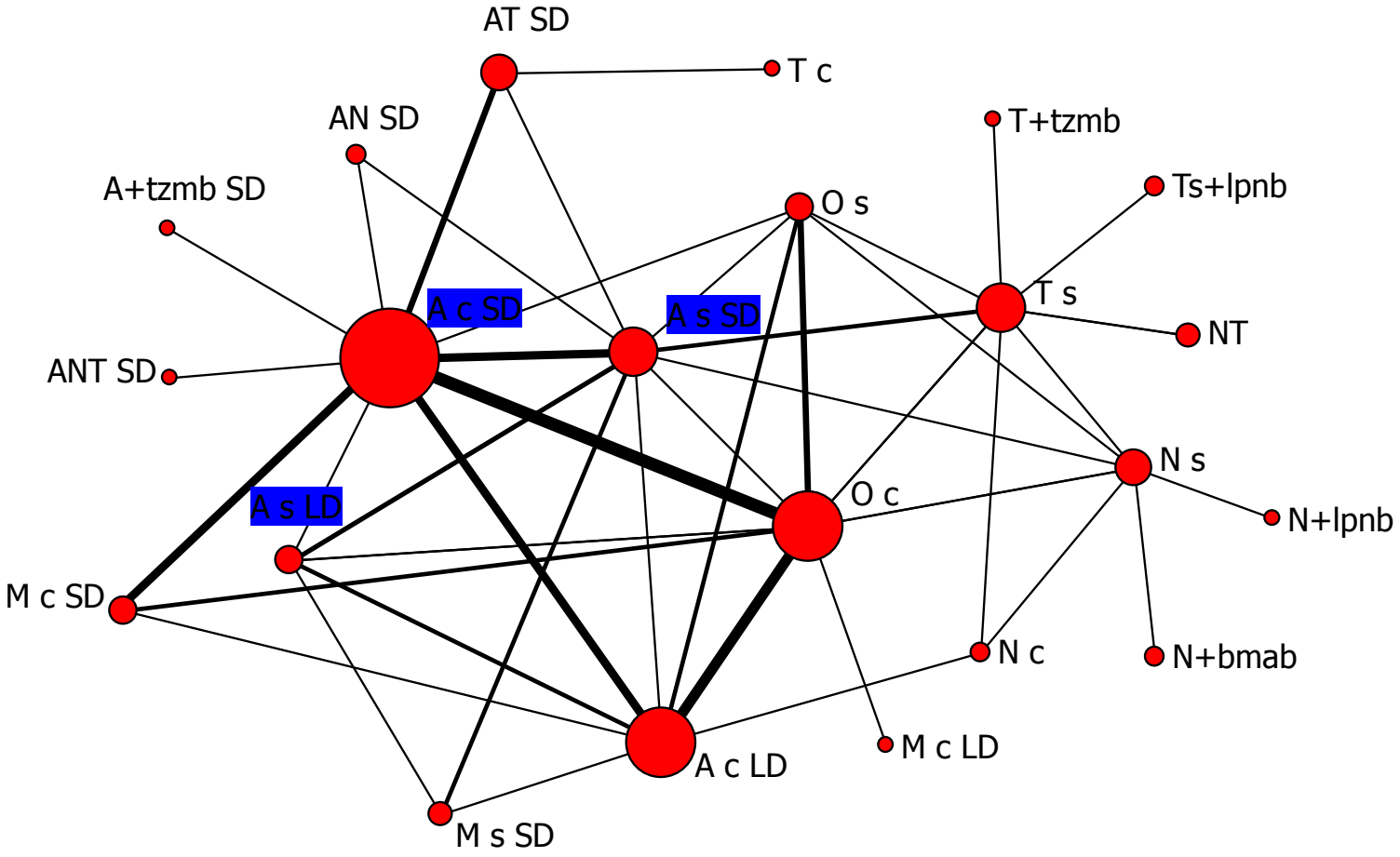
- It is stupid to depend on a single meta-analysis

- when there are many outcomes
- when there are many indications the same treatment comparison has been applied to
- when there are many other treatments and comparisons that have been considered for each of these indications

# Network definition

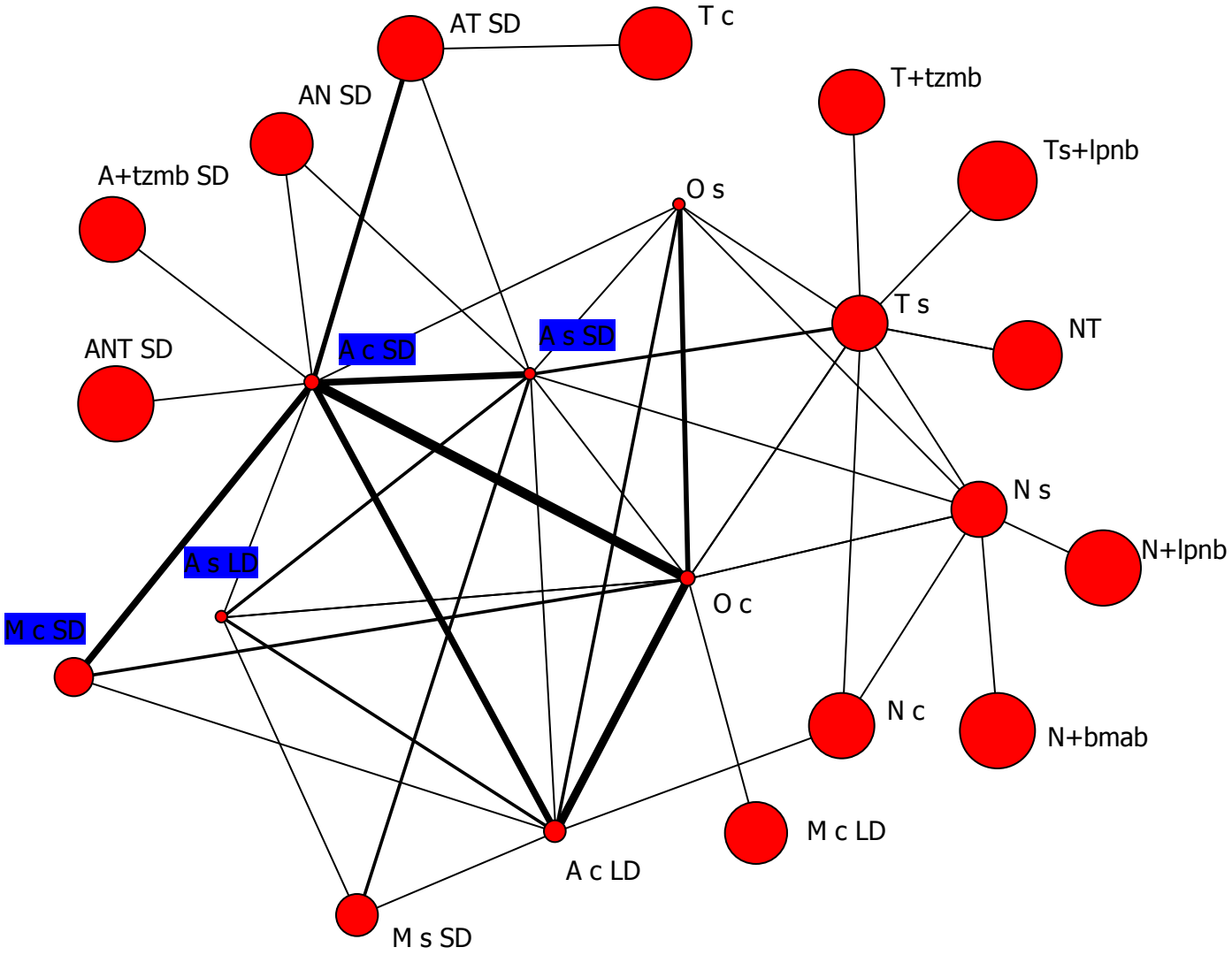
- Diverse pieces of data that pertain to research questions that belong to a wider agenda
- Information on one research question may indirectly affect also evidence on and inferences from other research questions
- In the typical application, data come from trials on different comparisons of different interventions, where many interventions are available to compare

# A network offers a wider picture than a single traditional meta-analysis: e.g. making sense of 700 trials of advanced breast cancer treatment




Mauri et al, JNCI 2008

# Focusing on what is most recent in the market



# Main types of network geometry

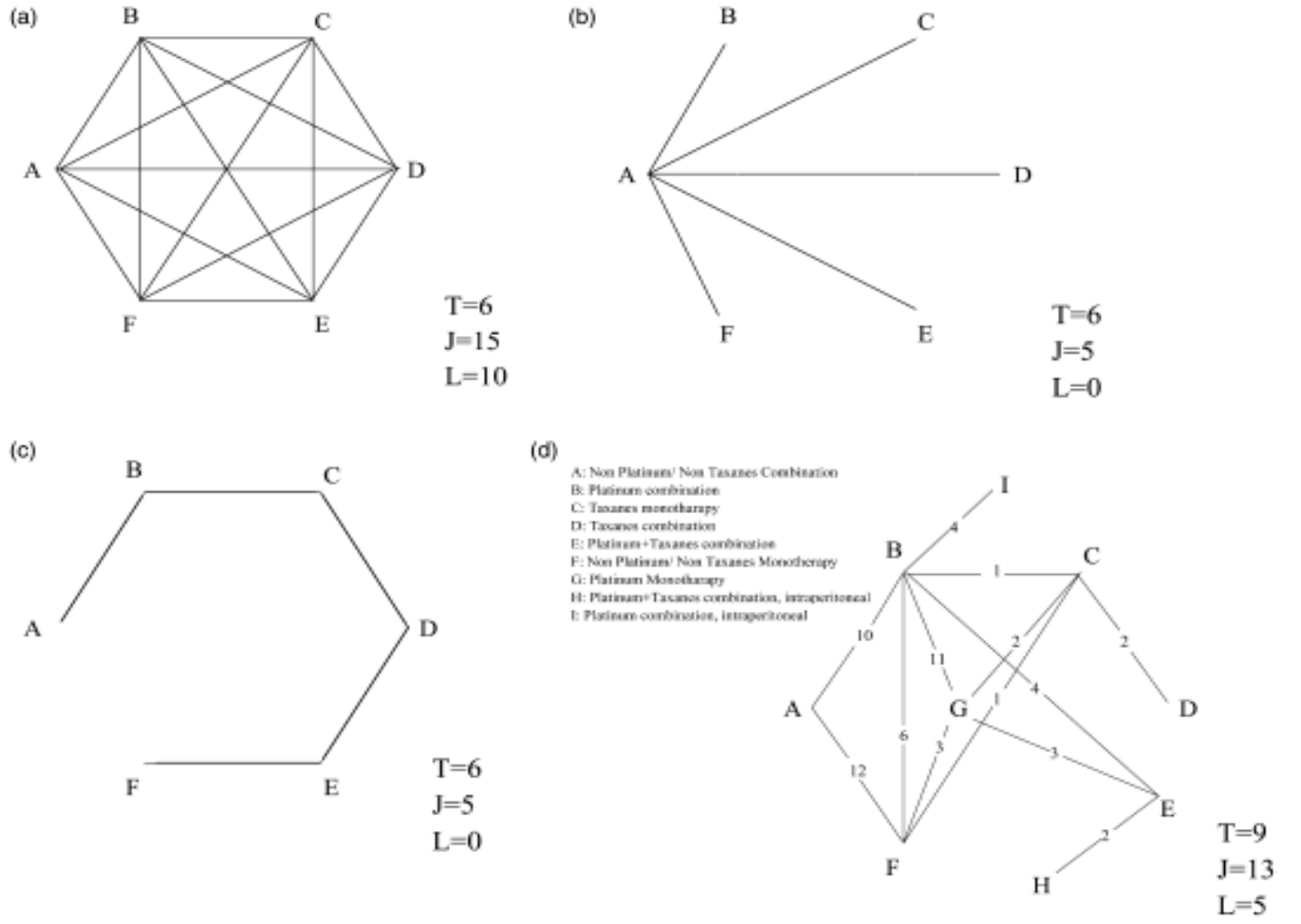



**Figure 2** Different geometries for a network: (a) full connected polygon, (b) radiating star, (c) linear structure, and (d) mixed example from real data.[9] For the last panel, each line connecting two treatments also shows the number of available direct comparisons between the two treatments.

*Polygons*

*Stars*

*Lines*

*Complex figures*

Salanti, Higgins, Ades, Ioannidis, Stat Methods Med Res 2008

# Diversity and co-occurrence

- Diversity = how many treatments are available and have they been equally well studied
- Co-occurrence = is there preference or avoidance for comparisons between specific treatments

Salanti, Kavvoura, Ioannidis, Ann Intern Med 2008

# Diversity: PIE (probability of interspecific encounter = probability that two randomly selected treatment groups (without replacement) belong to two different treatments)

## Properties of PIE and PIE′

Consider a network that refers to $N$ treatments that we are want to rank, for example, according to their efficacy. These treatments have been tested in several randomized trials, each with at least 2 groups. Let $S$ denote the total number of groups from all included studies featured in the network. If $p_i$ represents the proportion of the total number of groups in which treatment $i$ was tested, then:

$$PIE = \left( \frac{S}{S-1} \right) \left( 1 - \sum_{i=1}^{N} P_i^2 \right) \quad (1)$$

The PIE index is easily interpreted as a probability and is one of the few indices that is unbiased by sample size, although the variance increases if $S$ is small. PIE takes values in the interval:

$$PIE = \left[ \frac{(N\text{-}1)(2S-N)}{S(S-1)}, \frac{S(N-1)}{N(S-1)} \right] \quad (2)$$

$$PIE' = \frac{PIE}{PIE_{max}}$$

PIEmax varies according to the number of studies, e.g.
0.818 with 6 studies,
0.771 with 18 studies,
0.761 with 36 studies

**Properties of Co-occurrence Metrics**

In ecology, given a set of $N$ species $T = (T_1, T_2, \ldots, T_N)$ and some observations on their existence at $M$ different sites (imagine $M$ islands), one can estimate measures of the patterns of co-existence for the species. The data form an $N \times M$ matrix with entries 0 or 1 to denote the absence or presence of species on an island. Here, the different species are treatments and the islands are the studies. Co-occurrence demonstrates a pattern of favorite couples, in which treatment $X$ is usually compared to treatment $Y$ rather than $Z$, and $Z$ is usually compared with $W$ rather than anything else.

An important measure of co-occurrence (or lack thereof) is the checkerboard units for a particular pair of treatments. For 2 treatments $X$ and $Y$ that appear $Q_X$ and $Q_Y$ times in the network, and denoting the number of studies that directly compare $X$ versus $Y$ as $Q_{XY}$, we can calculate the number of checkerboard units ($CU$) as:

$$\mathrm{CU}_{XY} = (Q_X - Q_{XY})(Q_Y - Q_{XY})$$

This is the number of all possible pairs of studies that contain 1 treatment but not the other, and therefore measures the lack of co-occurrence. To evaluate the whole network of $N$ treatments, the number of checkerboard units is simply:

$$\sum_{\mathrm{X,Y}} CU_{XY} \tag{4}$$

that is, the sum of all checkerboard units for each pair of treatments. This calculates the total number of times that treatment pairs do not co-occur.

The C-score is obtained by averaging over all possible pairs of the N treatments:

$$C\text{-}score = \frac{\sum_{X,Y} CU_{XY}}{\left(\frac{N}{2}\right)} \tag{5}$$

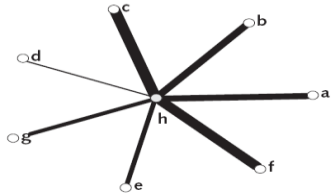
A larger C-score indicates a selective pattern in the choices of the compared treatments.

# Co-occurrence

## Checkerboard units

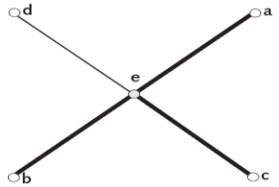
A

Second-generation antiepileptic drugs in partial epilepsy



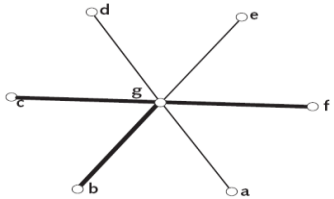

a: levetiracetam, b: gabapentin, c: lamotrigine, d: oxcarbazepine, e: tiagabine, f: topiramate, g: zonisamide, h: placebo

Biologic treatments for rheumatoid arthritis




a: adalimumab, b: infliximab, c: etanercept, d: anakinra, e: placebo

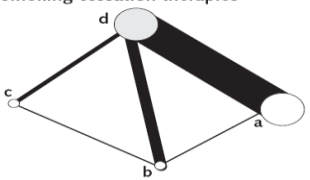
Intracoronary drug-eluting stents vs. bare-metal stents




a: AES, b: apolymeric PES, c: polymeric SES, d: MES, e: polymeric EES, f: polymeric PES, g: BMS

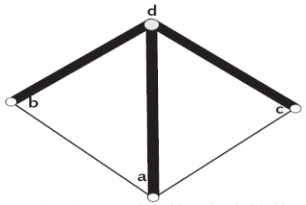
B

Smoking cessation therapies




a: nicotine replacement therapy, b: buproprion, c: varenicline, d: placebo/no treatment

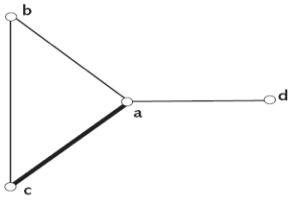
C Vitamin D and analogues to prevent bone loss and fractures




a: vit D/vitD + Ca, b: alfacalcidol/alfacalcidol + Ca, c: calcitriol/calcitriol + Ca, d: placebo/Ca

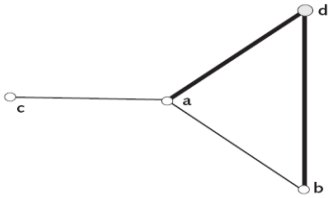
C *continued*

Self-monitoring of glucose in type 2 diabetes



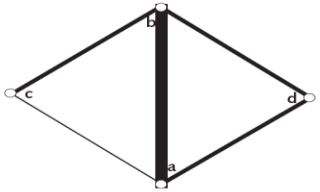

a: self-monitoring of blood glucose, b: self-monitoring of urine glucose, c: no self-monitoring, d: self-monitoring of blood glucose with regular feedback

Antiretroviral resistance testing in treatment-experienced patients




a: PART, b: GART, c: vPART, d: empiric

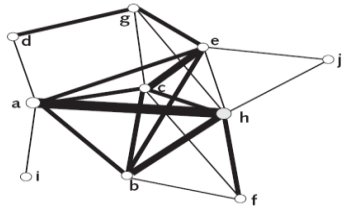
Prophylaxis for *Pneumocystis carinii* in HIV-infected patients




a: trimethoprim–sulfamethoxazole, b: pyrimethamine, c: dapsone, d: dapsone–pyrimethamine

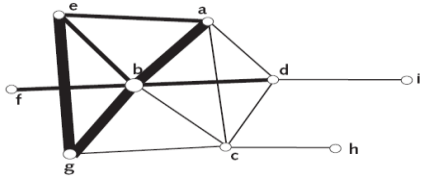
D

First-line antihypertensive therapy




a: diuretics, b: β-blockers, c: CCB, d: nonhydropiridine CCB, e: ACE-i, f: ARB, g: diuretics or β-blockers, h: placebo/not treated, i: α-blocker, j: ACE-i + diuretics

Chemotherapy regimens for ovarian cancer

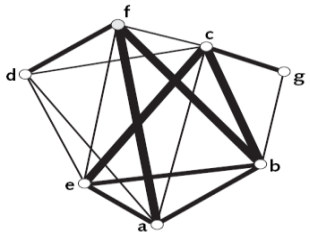



a: platinum monotherapy, b: platinum-based combination, c: taxane monotherapy, d: platinum + taxane-based combination, e: nonplatinum/nontaxane monotherapy, f: platinum-based combination (ip), g: nonplatinum/nontaxane combination, h: taxane-based combination, i: platinum/taxane-based combination (ip)

Shaded nodes indicate placebo or no active treatment. The thickness of the lines is proportional to the number of trials addressing each specific comparison. **A.** Star-shaped networks. **B.** Non–star-shaped networks with limited diversity and significant co-occurrence. **C.** Non–star-shaped networks with limited diversity and nonsignificant co-occurrence. **D.** Networks with considerable diversity and significant co-occurrence. **E.** Networks with considerable diversity and nonsignificant co-occurrence. ACE-I = angiotensin-converting enzyme

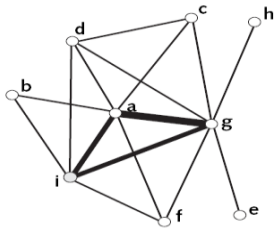
**E**

Antihypertensive treatment (incidence of diabetes)



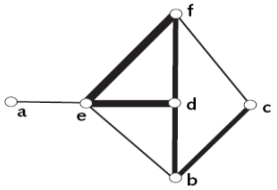

a: diuretic, b: ACE-i, c: CCB, d: ARB, e: β-blocker, f: placebo, g: β-blocker or diuretic

Stroke prevention in nonrheumatic atrial fibrillation




a: aspirin, b: alternate-day aspirin, c: fixed low-dose-warfarin, d: fixed low-dose warfarin and aspirin, e: indobufen, f: adjusted low-dose warfarin, g: adjusted standard-dose warfarin, h: ximelagatran, i: placebo/control

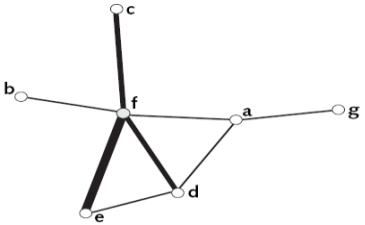
Treatments for acute myocardial infarction




a: anistreplase, b: accelerated t-PA, c: reteplase, d: angioplasty, e: streptokinase, f: t-PA

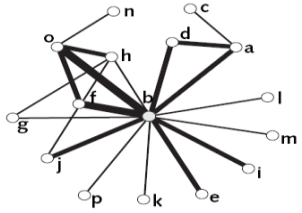
**F**

α-1 antagonists in lower urinary tract symptoms




a: alfuzocin, b: alfuzocin SR, c: doxazocin, d: tamsulocin, e: terazocin, f: placebo, g: prazosin

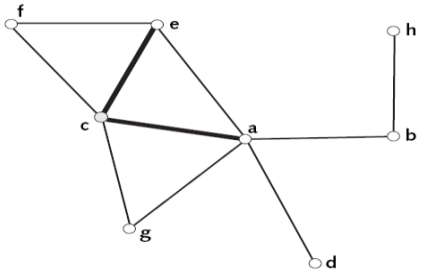
Topical nonsteroidal anti-inflammatory drugs for acute pain




a: indomethacin, b: placebo, c: indomethacin + placebo, d: piroxicam, e: niflumic acid, f: ibuprofen, g: ketorolac, h: etofenamate, i: diclofenac, j: felbinac, k: fentiazac, l: naproxen, m: meclofenamic acid, n: flunoxaprofen, o: ketoprofen, q: flurbiprofen

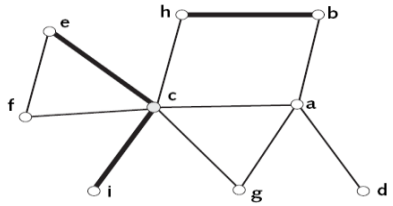
**G**

Antifungal prophylaxis in liver transplant recipients



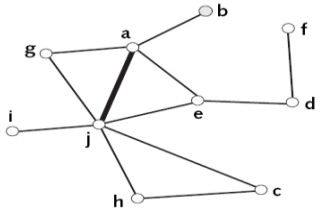

a: fluconazole, b: nystatin, c: placebo, d: clotrimazole, e: itraconazole, f: fluconazole + itraconazole, g: amphotericin, h: liposomal amphotericin B

Antifungal prophylaxis in solid organ transplant recipients




a: fluconazole, b: nystatin, c: placebo, d: amphotericin B, e: liposomal amphotericin B, f: fluconazole + itraconazole, g: itraconazole, h: clotrimazole, i: ketoconazole

Topical antibiotics without steroids for chronic ear discharge without eardrum perforation




a: ciprofloxacin, b: placebo, c: ofloxacin, d: TSP, e: gentamicin, f: TP, g: tobramycin, h: neomycin–polymyxin, i: chloramphenicol/gentamycin, j: antiseptic

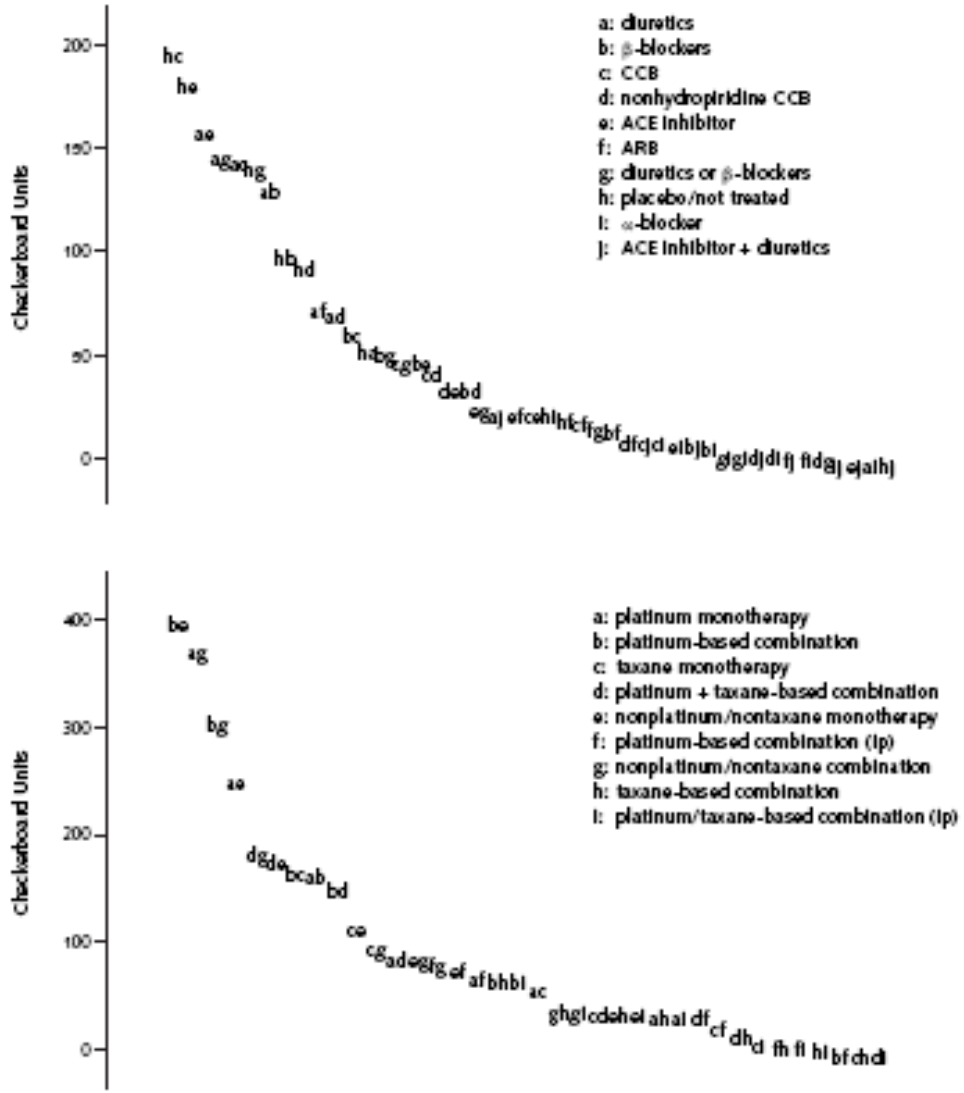
*Table.* **Geometry Metrics for the Treatment Networks***

| Network (Reference) | Treatments, *n* | Studies, *n* | Groups, *n* | PIE Index | PIE* Index | C-Score | *P* Value |
|---|---|---|---|---|---|---|---|
| Treatments for acute myocardial infarction (3) | 6 | 14 | 32 | 0.823 | 0.853 | 13.8 | 0.48 |
| Second-generation antiepileptic drugs in partial epilepsy (15) | 8 | 37 | 74 | 0.711 | 0.802 | 19.71 | NA† |
| Biologic treatments for rheumatoid arthritis (16) | 5 | 17 | 34 | 0.706 | 0.857 | 10.7 | NA† |
| Intracoronary drug-eluting stents vs. bare-metal stents (17) | 7 | 17 | 34 | 0.717 | 0.812 | 5.38 | NA† |
| Smoking cessation therapies (18) | 4 | 84 | 174 | 0.592 | 0.785 | 176.33 | <0.001 |
| Vitamin D and analogues to prevent bone loss and fractures (19) | 4 | 35 | 70 | 0.69 | 0.853 | 80 | 1.0 |
| Self-monitoring of glucose in type 2 diabetes (20) | 4 | 13 | 27 | 0.664 | 0.853 | 6.67 | 1.0 |
| Antiretroviral resistance testing (21) | 4 | 10 | 21 | 0.743 | 0.943 | 9 | 0.056 |
| Prophylaxis for *Pneumocystis carinii* in HIV (22) | 4 | 22 | 48 | 0.71 | 0.927 | 30 | 0.83 |
| First-line antihypertensive therapy (23) | 10 | 43 | 93 | 0.83 | 0.912 | 46.96 | <0.001 |
| Chemotherapy regimens for ovarian cancer (24) | 9 | 58 | 117 | 0.822 | 0.917 | 95.53 | <0.001 |
| Antihypertensive treatment (incidence of diabetes) (25) | 7 | 22 | 48 | 0.858 | 0.98 | 27.33 | 0.88 |
| Stroke prevention in nonrheumatic atrial fibrillation (26) | 9 | 19 | 45 | 0.828 | 0.911 | 9.78 | 0.92 |
| $\alpha_1$-Antagonists in lower urinary tract symptoms (27) | 7 | 25 | 50 | 0.759 | 0.868 | 19.57 | 0.29 |
| Topical nonsteroidal anti-inflammatory drugs for acute pain (28) | 16 | 34 | 72 | 0.822 | 0.865 | 9.38 | 0.41 |
| Antifungal prophylaxis in liver transplantation (29) | 8 | 10 | 21 | 0.857 | 0.933 | 3.78 | 0.64 |
| Antifungal prophylaxis in solid organ transplantation (30) | 9 | 14 | 29 | 0.859 | 0.945 | 6.17 | 0.24 |
| Topical antibiotics for chronic ear discharge (31) | 10 | 13 | 27 | 0.884 | 0.949 | 4.89 | 0.084 |

* HIV = human immunodeficiency virus; NA = not applicable; PIE = probability of interspecific encounter; PIE = PIE divided by the maximum value PIE can take for the given number of studies.

† *P* value could not be calculated because the network was fully star-shaped.

*Figure 2.* Distribution of checkerboard units for all possible comparisons.



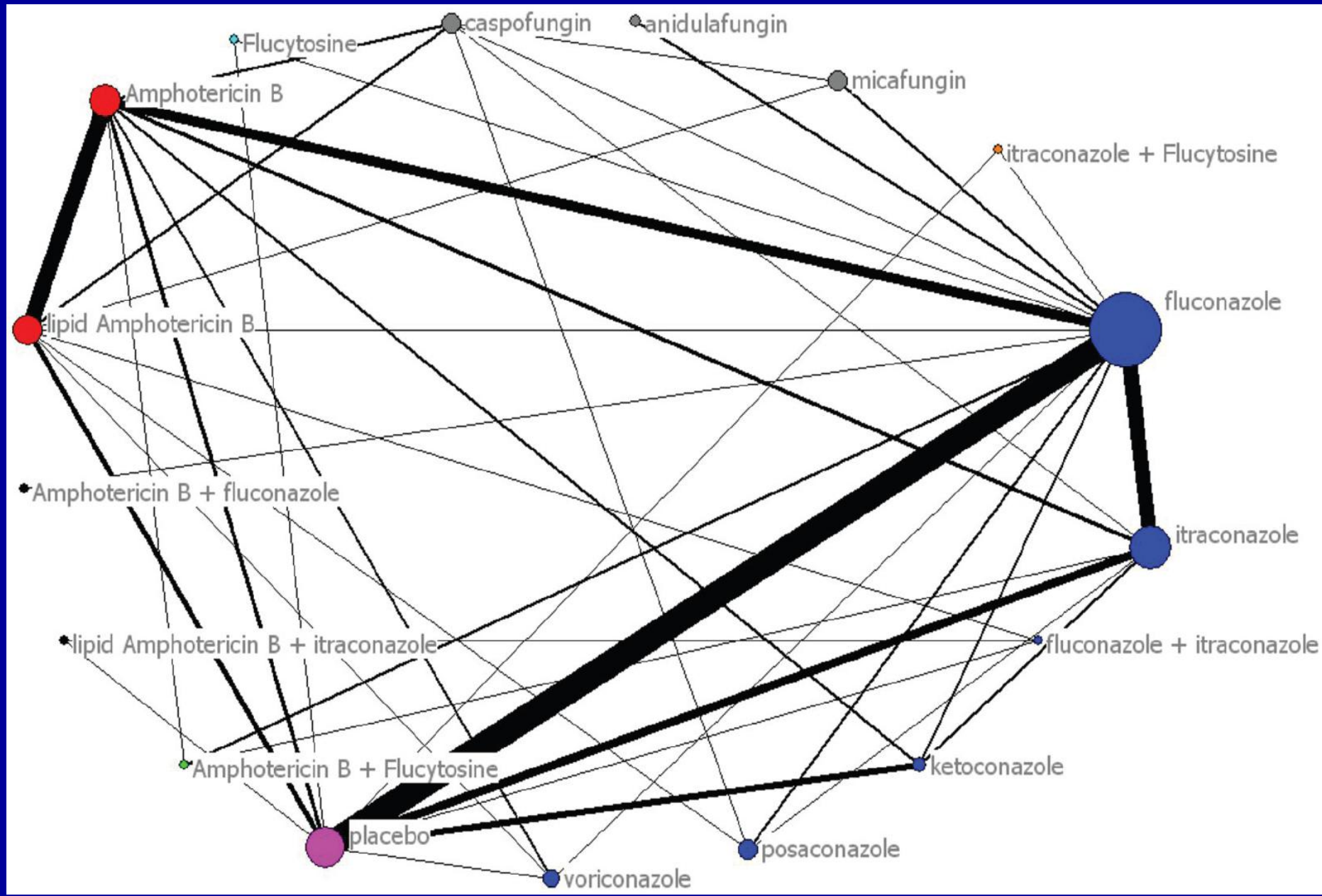

Treatment comparisons are shown in order of decreasing checkerboard units. Top. First-line antihypertensive therapy. Bottom. Chemotherapy for ovarian cancer.

# Homophily

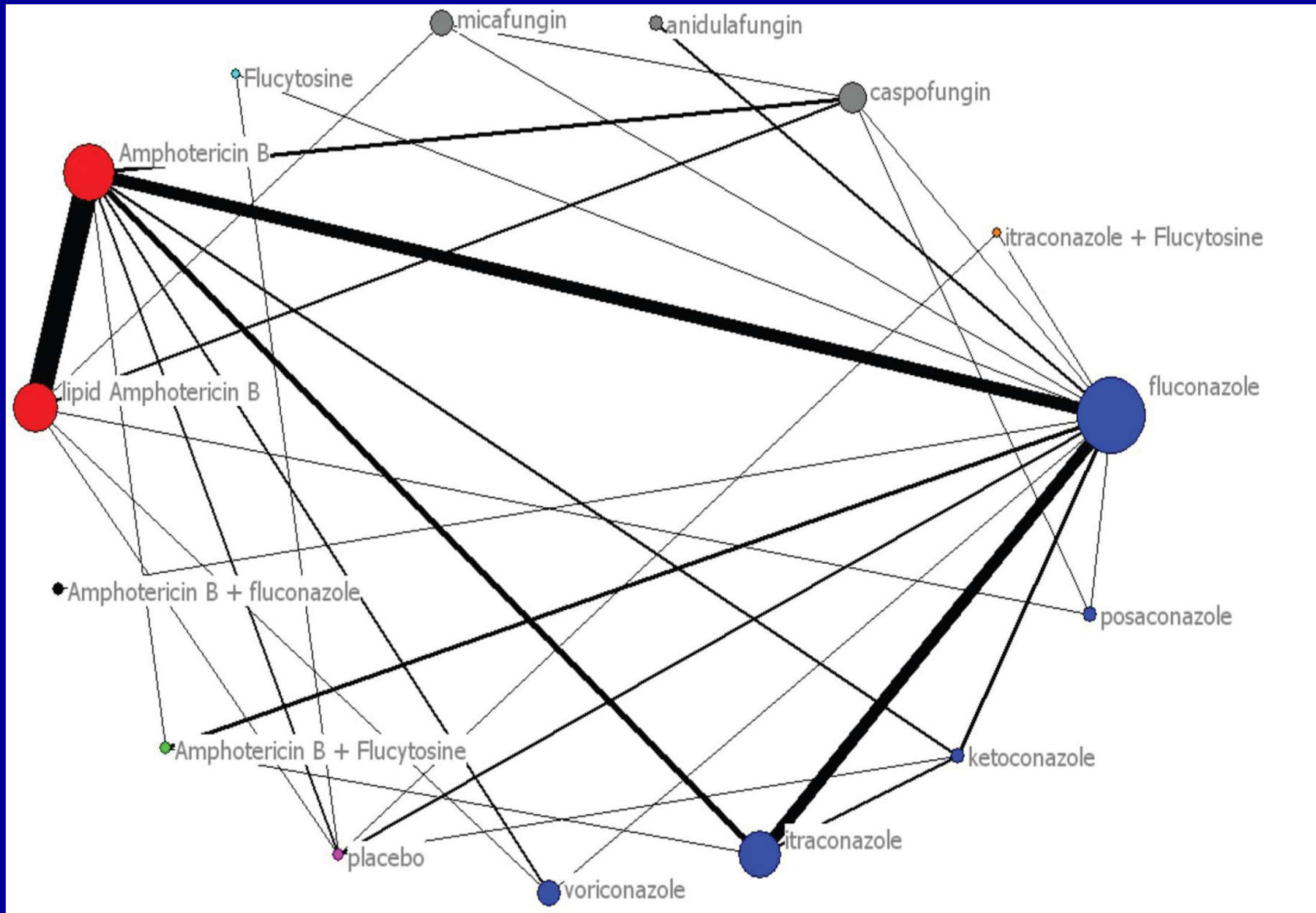
- ΟΜΟΦΙΛΙΑ = Greek for “love of the same” = birds of a feather flock together
- Testing for homophily examines whether agents in the same class are disproportionately more likely to be compared against each other than with agents of other classes.

# For example: Antifungal agents agenda

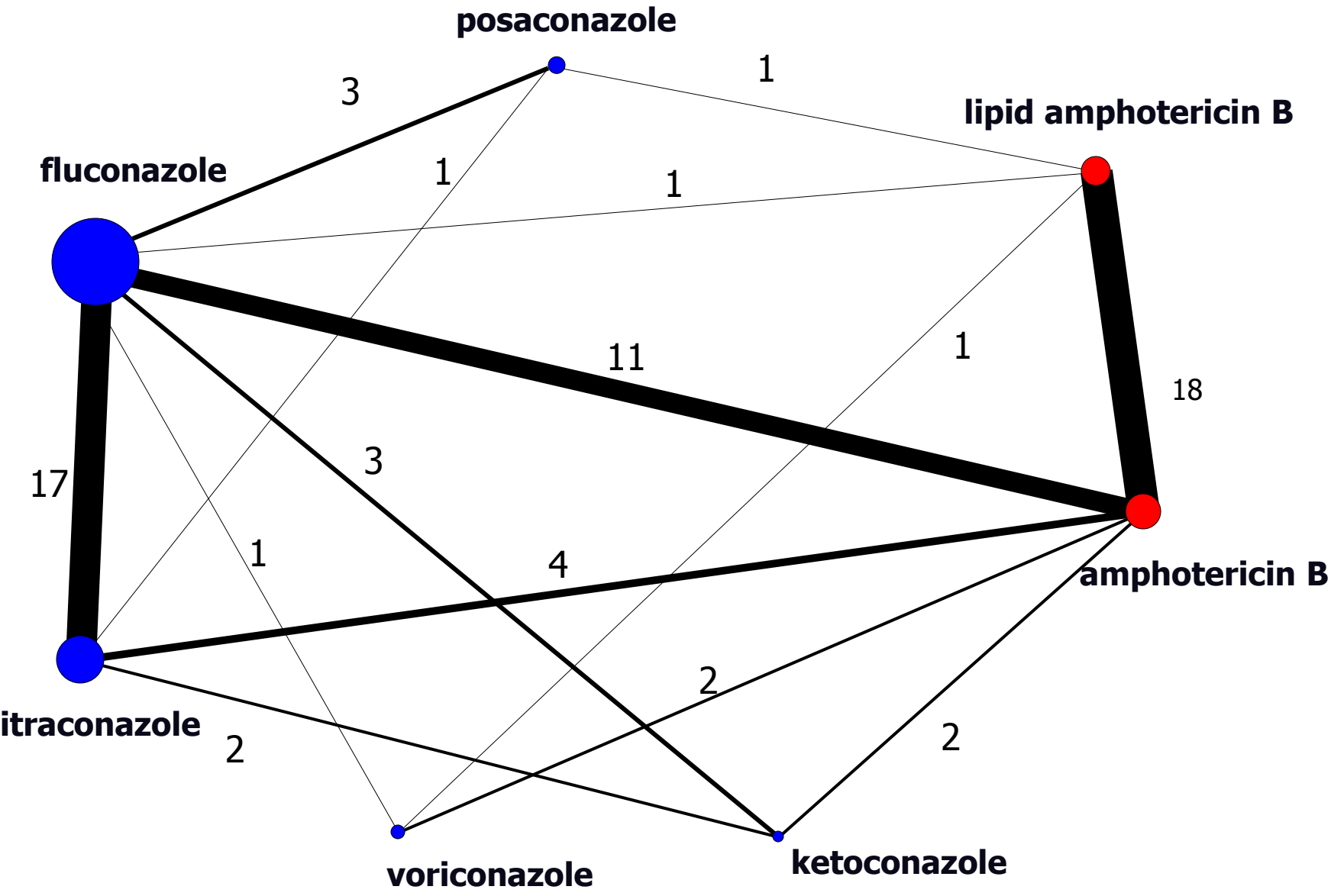
- Old classes: polyenes, old azoles
- New classes: echinocandins, newer azoles



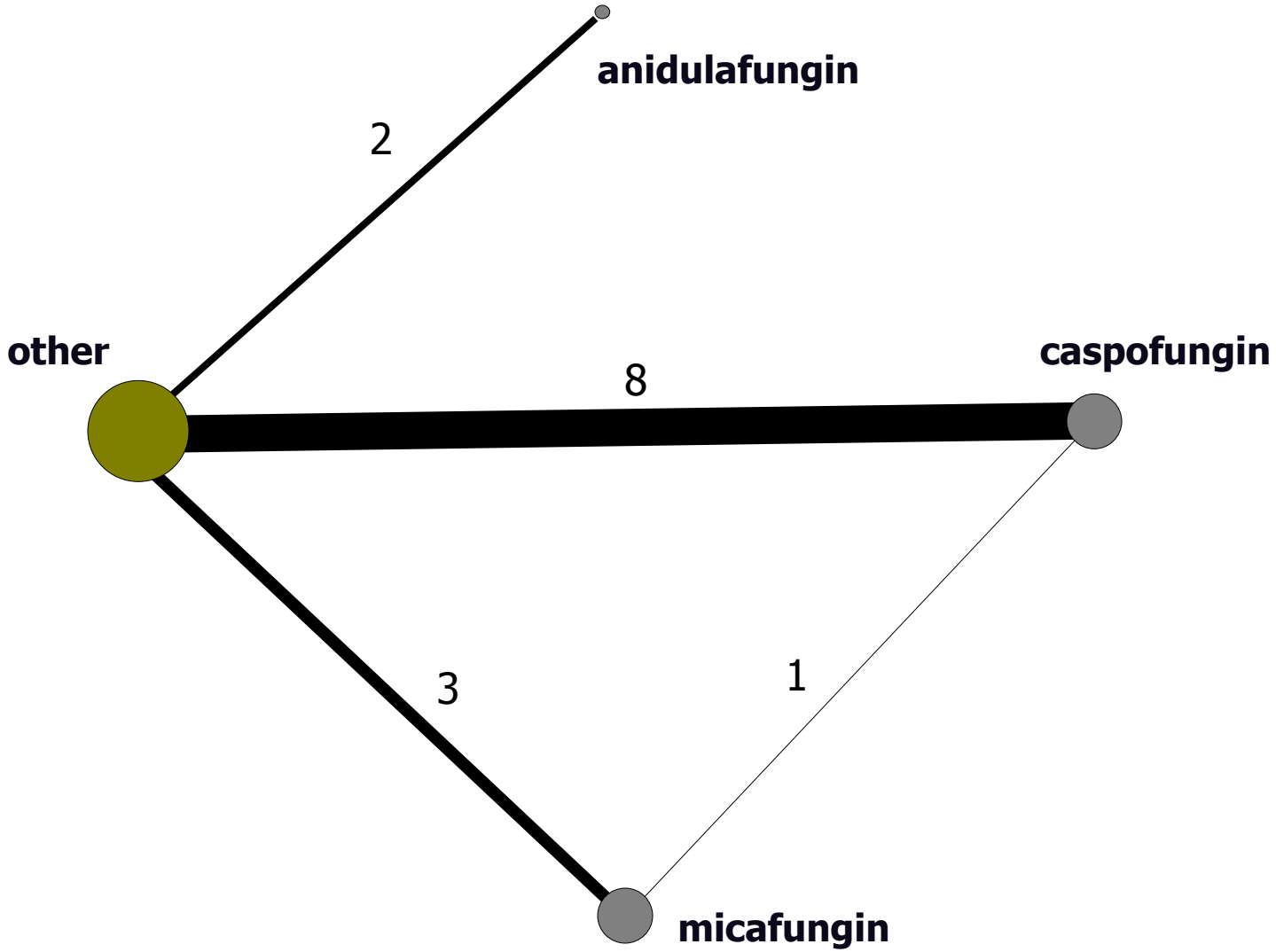

Rizos et al, J Clin Epidemiol, 2010

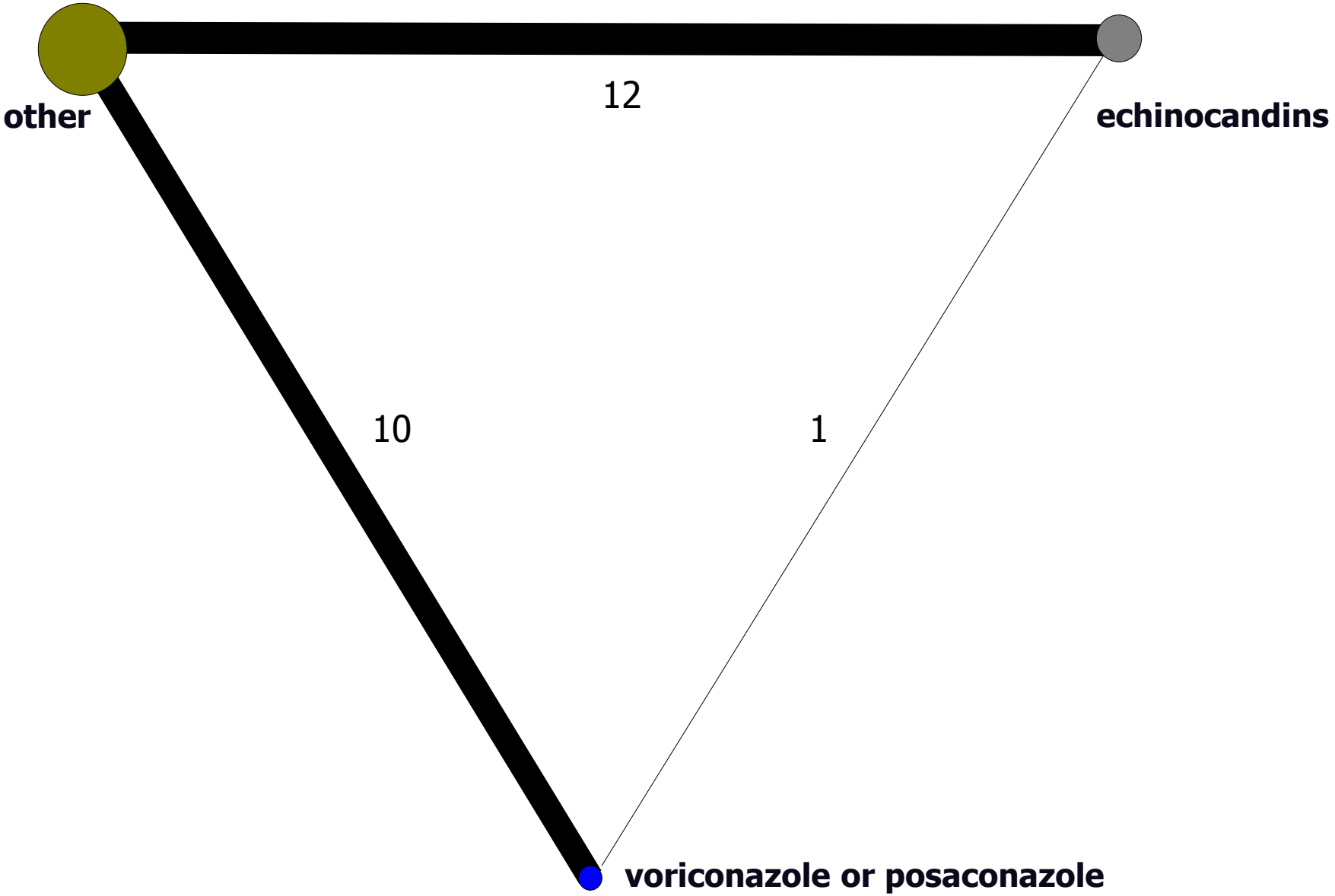

micafungin
anidulafungin
Flucytosine
caspofungin
Amphotericin B
itraconazole + Flucytosine
lipid Amphotericin B
fluconazole
Amphotericin B + fluconazole
posaconazole
Amphotericin B + Flucytosine
ketoconazole
itraconazole
placebo
voriconazole

- Among polyene and azole groups, agents were compared within the same class more often than they did across classes (homophily test $p<0.001$ for all trials).
- Lipid forms of amphotericin B were compared almost entirely against conventional amphotericin formulations (n=18 trials), with only 4 comparisons against azoles.


posaconazole
lipid amphotericin B
fluconazole
amphotericin B
itraconazole
voriconazole
ketoconazole
3
1
1
1
11
1
18
17
3
1
4
2
2
2

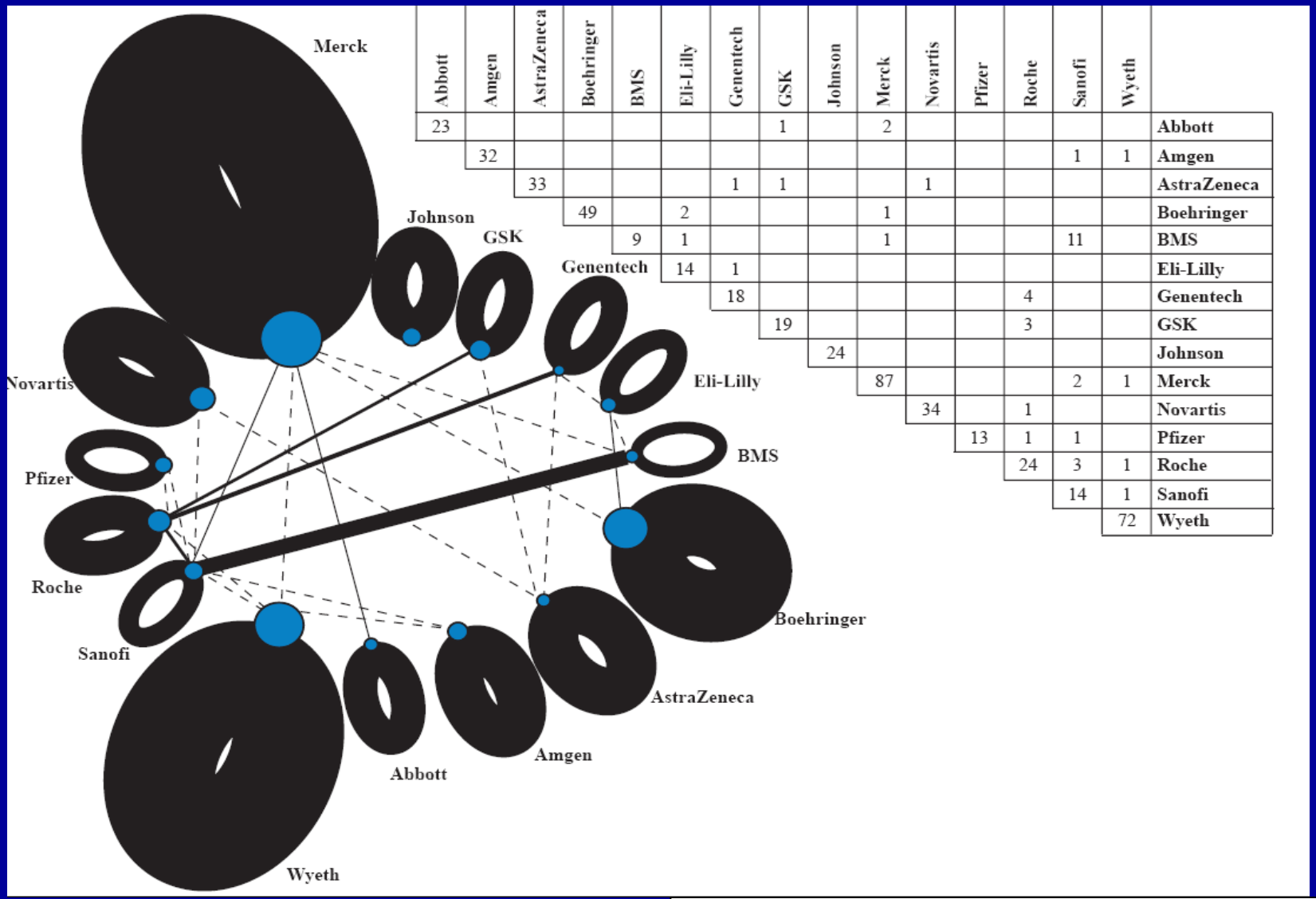
- There was strong evidence of avoidance of head-to-head comparisons for newer agents. Only one among 14 trials for echinocandins has compared head-to-head two different echinocandins ($p<0.001$ for co-occurrence). Of 11 trials on newer azoles, only one compared a newer azole with an echinocandin ($p<0.001$ for co-occurrence).


anidulafungin
2
other
caspofungin
8
3
1
micafungin


other
12
echinocandins
10
1
voriconazole or posaconazole

# Auto-looping

Design of clinical research: an open world or isolated city-states (company-states)?




| | Abbott | Amgen | AstraZeneca | Boehringer | BMS | Eli-Lilly | Genentech | GSK | Johnson | Merck | Novartis | Pfizer | Roche | Sanofi | Wyeth |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Abbott | 23 | | | | | | | 1 | | 2 | | | | | |
| Amgen | | 32 | | | | | | | | | | | | 1 | 1 |
| AstraZeneca | | | 33 | | | | 1 | 1 | | | 1 | | | | |
| Boehringer | | | | 49 | | 2 | | | | 1 | | | | | |
| BMS | | | | | 9 | 1 | | | | 1 | | | | 11 | |
| Eli-Lilly | | | | | | 14 | 1 | | | | | | | | |
| Genentech | | | | | | | 18 | | | | | | 4 | | |
| GSK | | | | | | | | 19 | | | | | 3 | | |
| Johnson | | | | | | | | | 24 | | | | | | |
| Merck | | | | | | | | | | 87 | | | | 2 | 1 |
| Novartis | | | | | | | | | | | 34 | | 1 | | |
| Pfizer | | | | | | | | | | | | 13 | 1 | 1 | |
| Roche | | | | | | | | | | | | | 24 | 3 | 1 |
| Sanofi | | | | | | | | | | | | | | 14 | 1 |
| Wyeth | | | | | | | | | | | | | | | 72 |

Lathyris et al., Eur J Clin Invest, 2010

# Synthesis of the network evidence (multiple-treatment meta-analysis)

- Incoherence
- Summary effects
- Ranking
- Bias modeling

# Credible intervals and predictive intervals in network meta-analysis

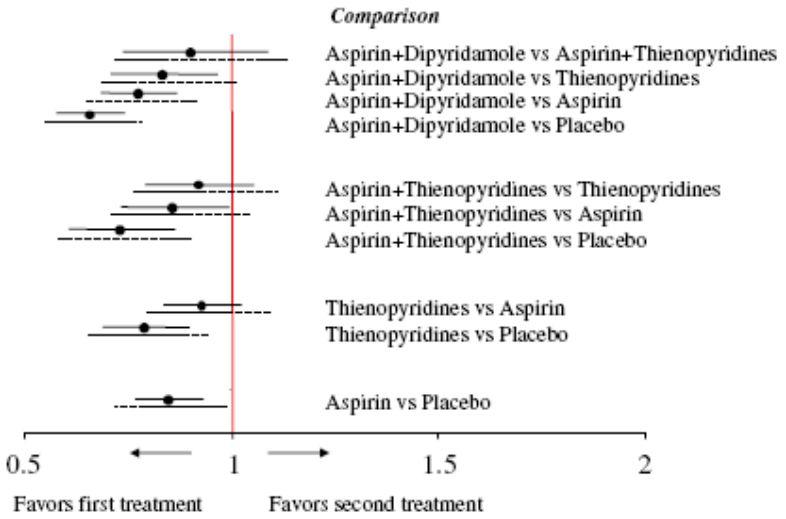



Fig. 1. Pairwise odds ratios and their 95% credible intervals (solid lines) for serious vascular events with antiplatelet treatments estimated with multiple-treatment meta-analysis. The dotted lines present the predictive intervals.

Salanti, Ades, Ioannidis, JCE, 2011

# Posterior distributions of effects and corresponding predictive distributions of effects

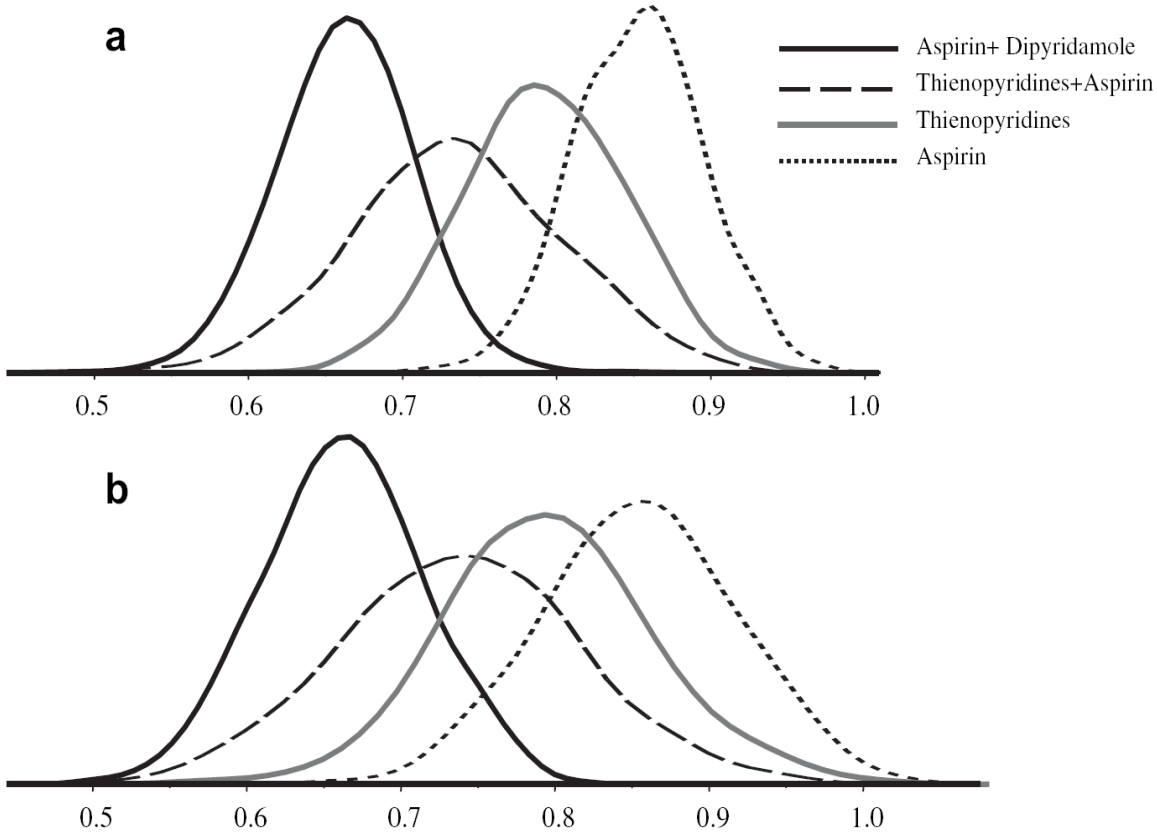
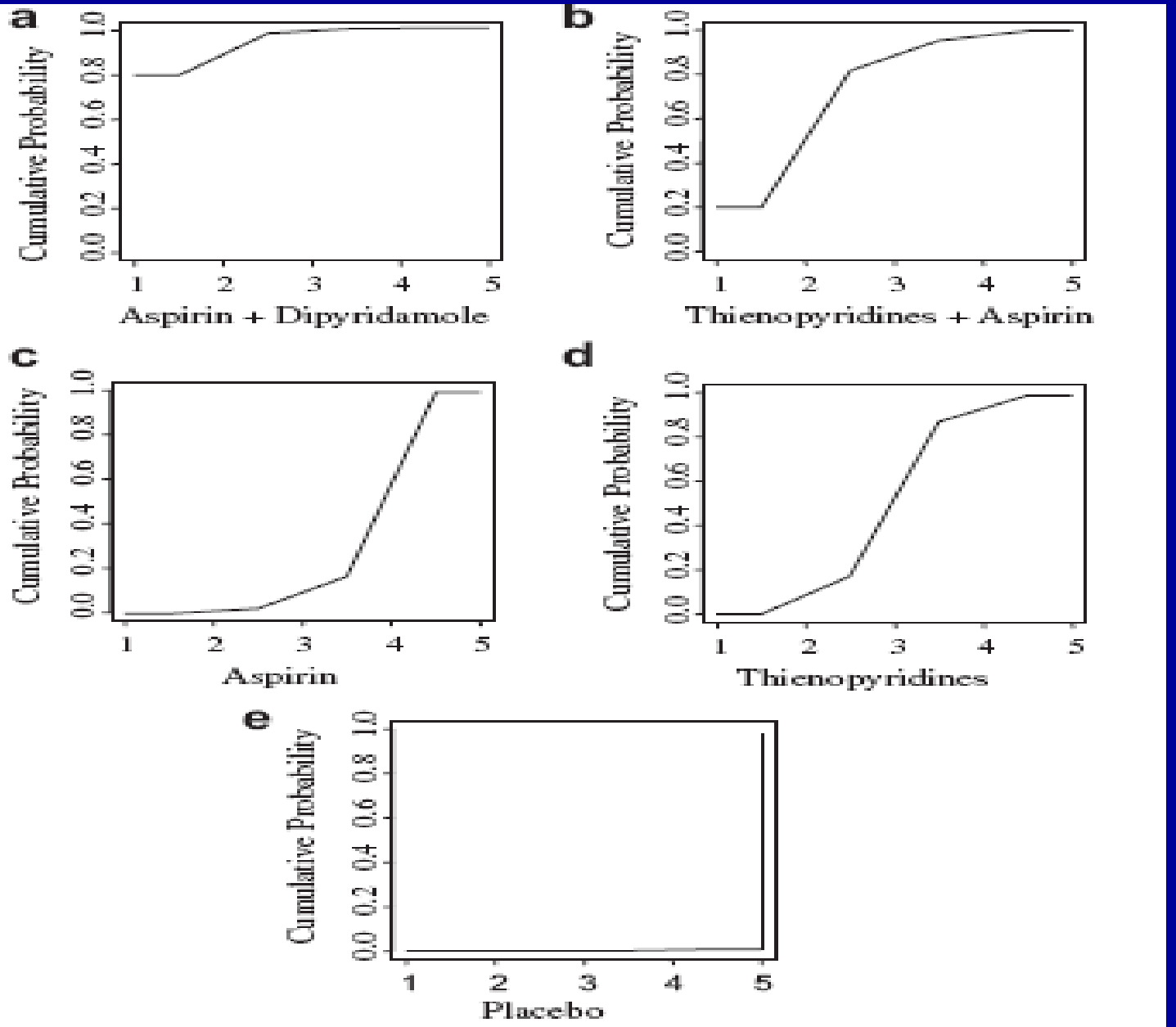



Fig. 2. Posterior distributions of mean odds ratios for serious vascular events with antiplatelet treatments compared with placebo (a) and the corresponding predictive distributions of effects within which the effect size of a new study is expected to be found with 95% probability (b).

JCE, 2011

# Cumulative ranking probability



# Probability of not being worse than threshold t from the best treatment

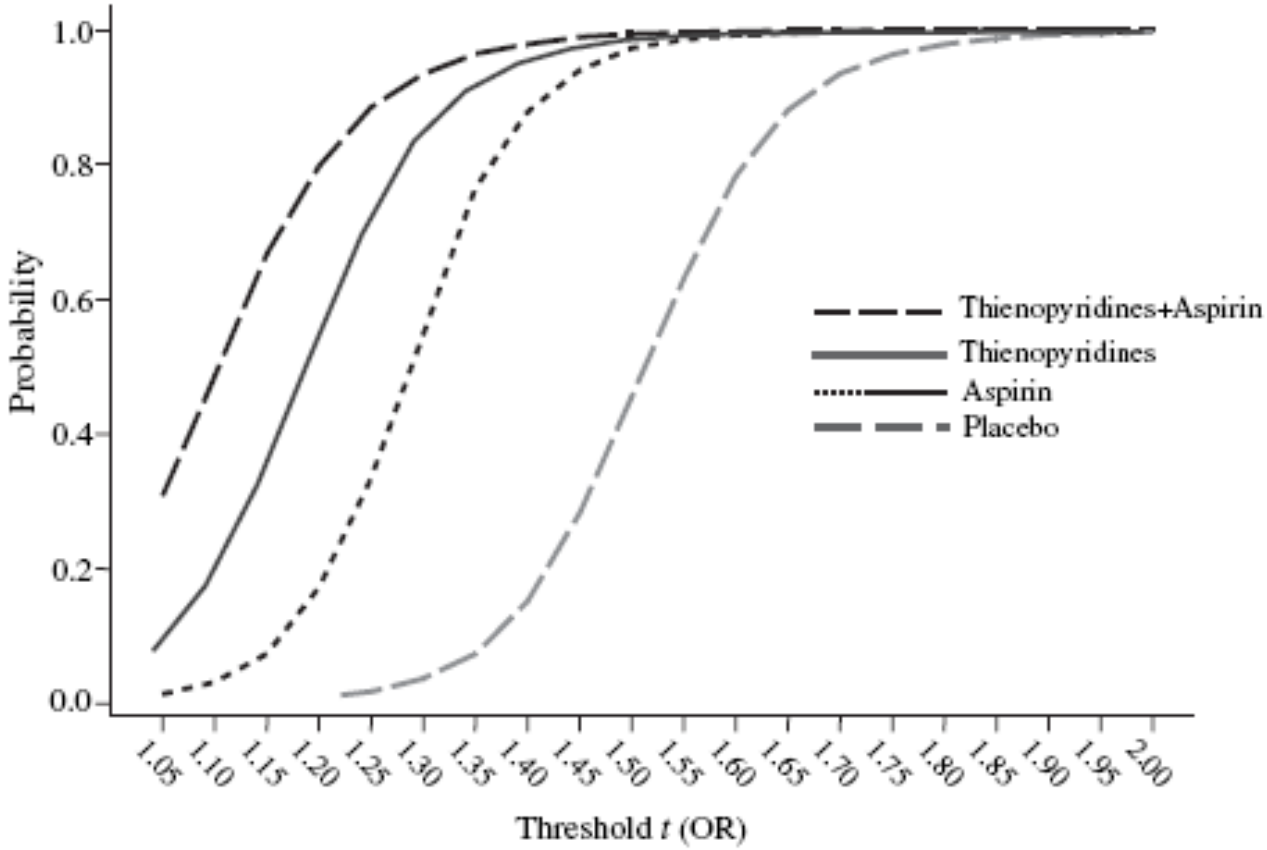



Fig. 6. Probabilities for each treatment to be no worse than the combination of aspirin and dipyridamole regarding the incidence of serious vascular events by a certain threshold *t* (on the horizontal axis) measured in odds ratio scale.

# Modeling bias

Research Article

Statistics in Medicine

Received 15 December 2009, Accepted 26 May 2010 View this article online at wileyonlinelibrary.com

(wileyonlinelibrary.com) DOI: 10.1002/sim.4001

## Evaluating novel agent effects in multiple-treatments meta-regression

Georgia Salanti,[a*†] Sofia Dias,[b] Nicky J. Welton,[b] AE Ades,[b] Vassilis Golfinopoulos,[c] Maria Kyrgiou,[d] Davide Mauri[a,e] and John P. A. Ioannidis[a,f,g]

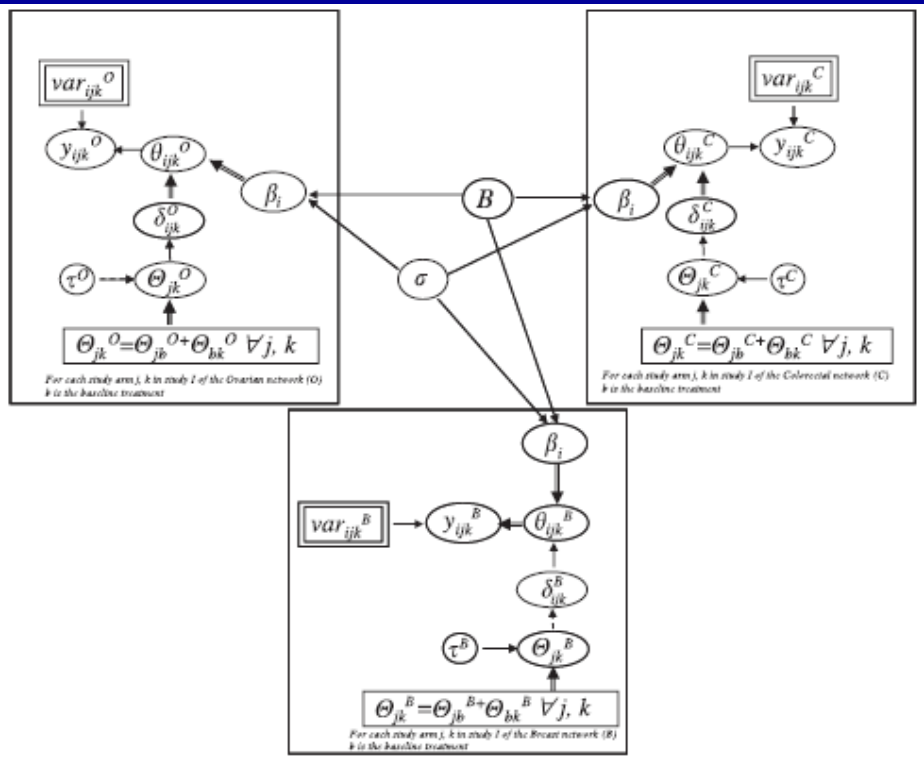



**Figure 2**. Graphical representation of the jointly adjusted model. Stochastic nodes (associated with distributions) and deterministic nodes (logical functions of parameters) are presented in oval shapes and data are presented in rectangular shapes. Single-line arrows represent distributions and double-line arrows represent logical functions.

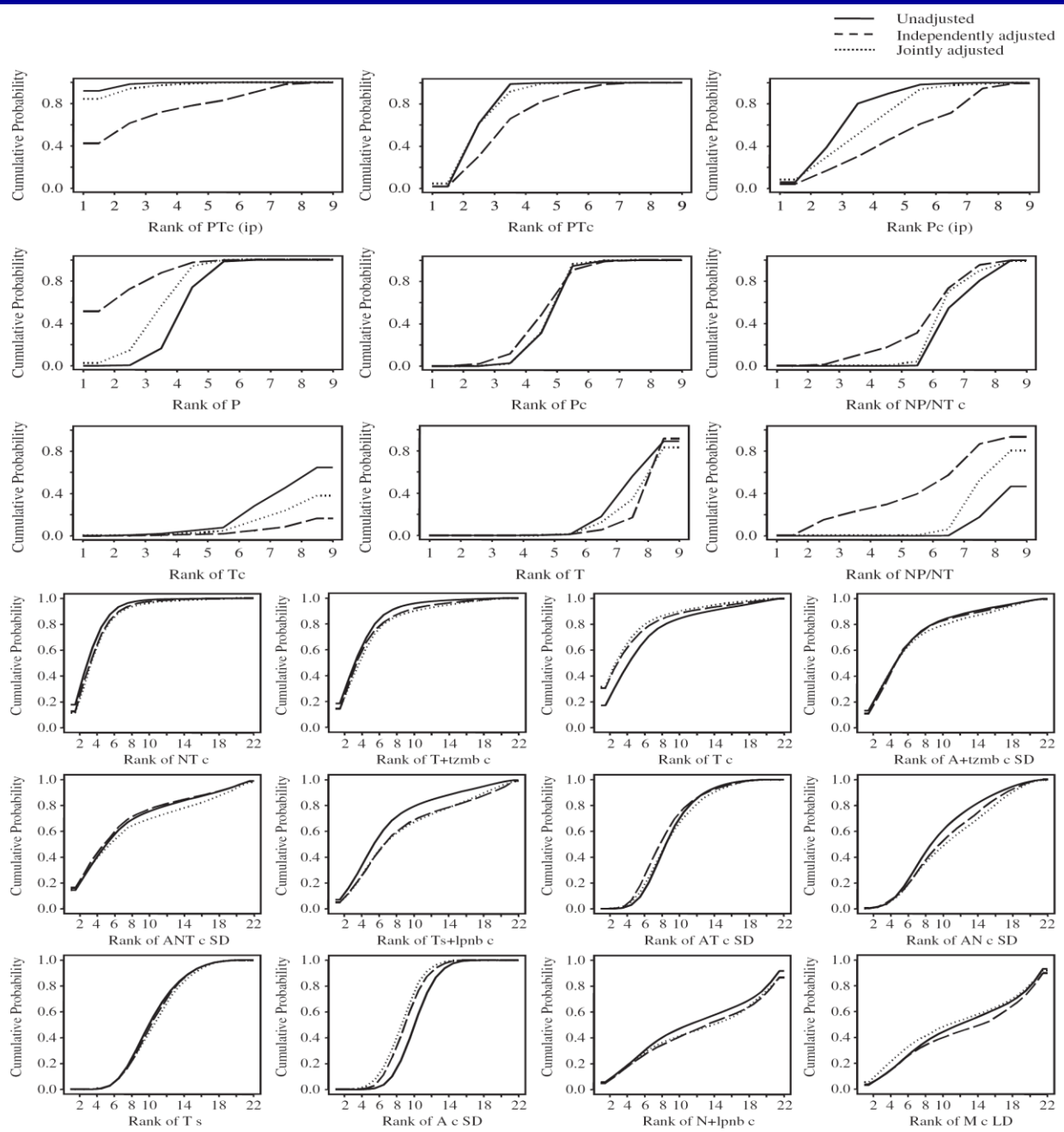



**Figure A1**. Cumulative ranking curves for the ovarian and breast cancer treatments obtained from the unadjusted model (solid line), the independently adjusted (dashed line) and jointly adjusted model (dotted line). The surface under each cumulative curves expressed as percentage is presented in Appendix Table AII.

# Changes in cumulative ranking

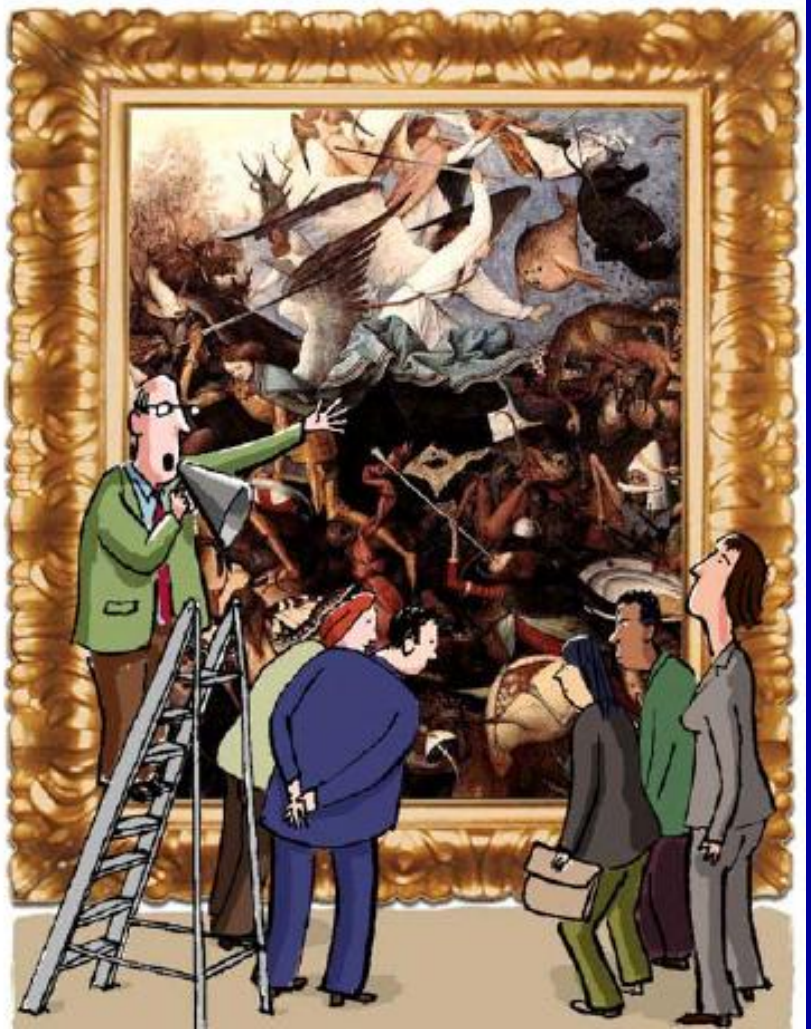
# Reversing the paradigm

Design networks prospectively

- Data are incorporated prospectively
- Geometry of the research agenda is pre-designed
- Next study is designed based on enhancing, improving geometry of the network, and maximizing the informativity given the network

# The need to consider the wider agenda in systematic reviews and meta-analyses

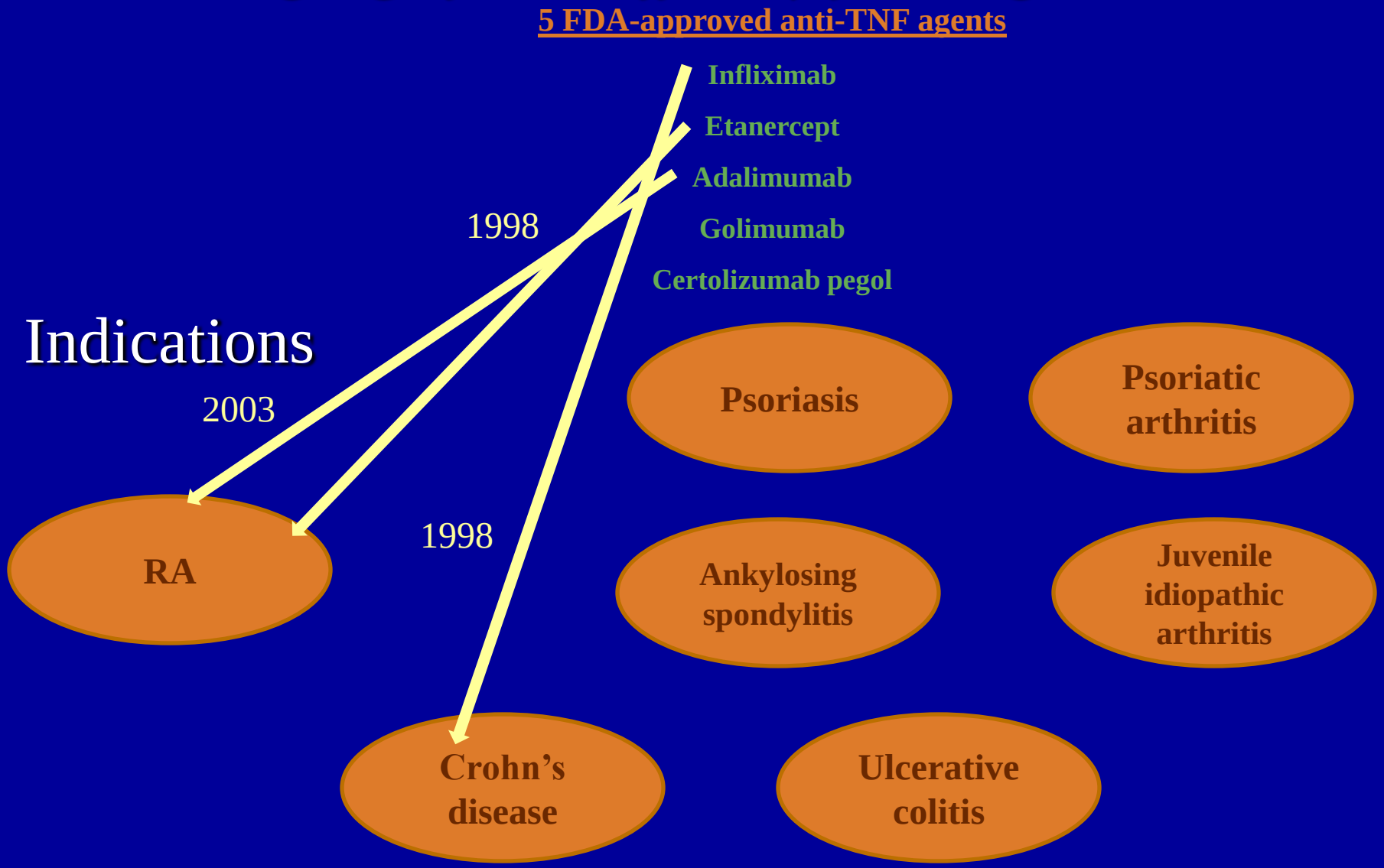
As well as focusing on a precise question, systematic reviewers also need to consider the whole research programme for the interventions under study, argue **John Ioannidis** and **Fotini Karassa**



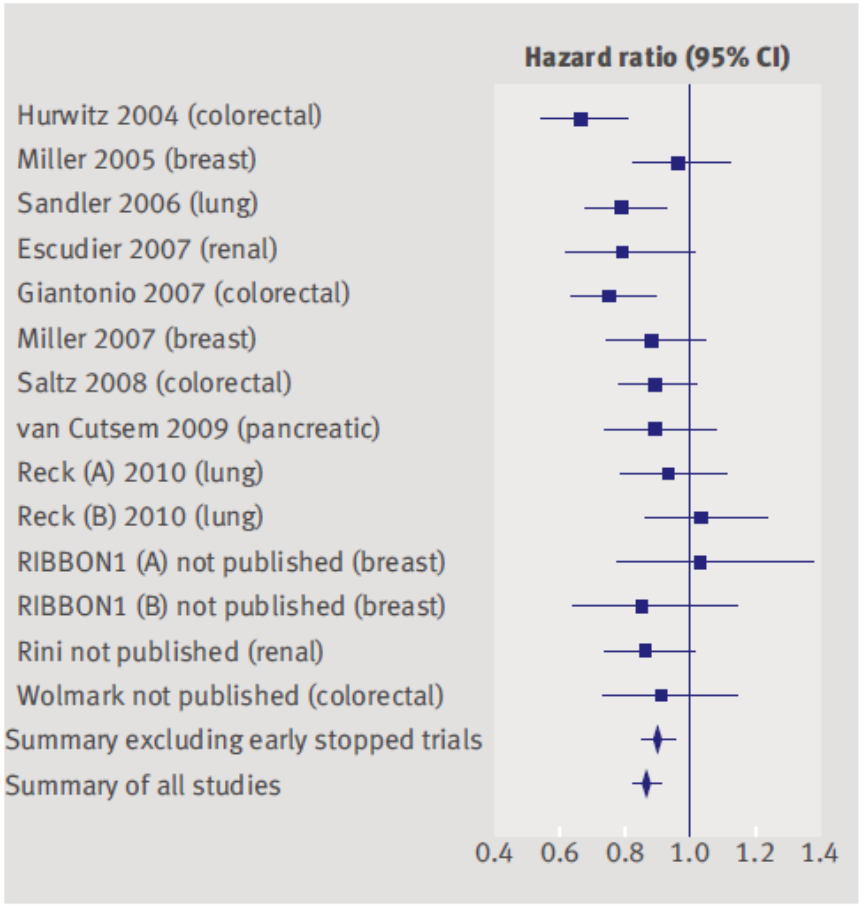
This may be happening already?

Agenda-wide meta-analyses
BMJ 2010

# *Anti-TNF agents: $ 10 billion and 43 meta-analyses, all showing significant efficacy for single indications*

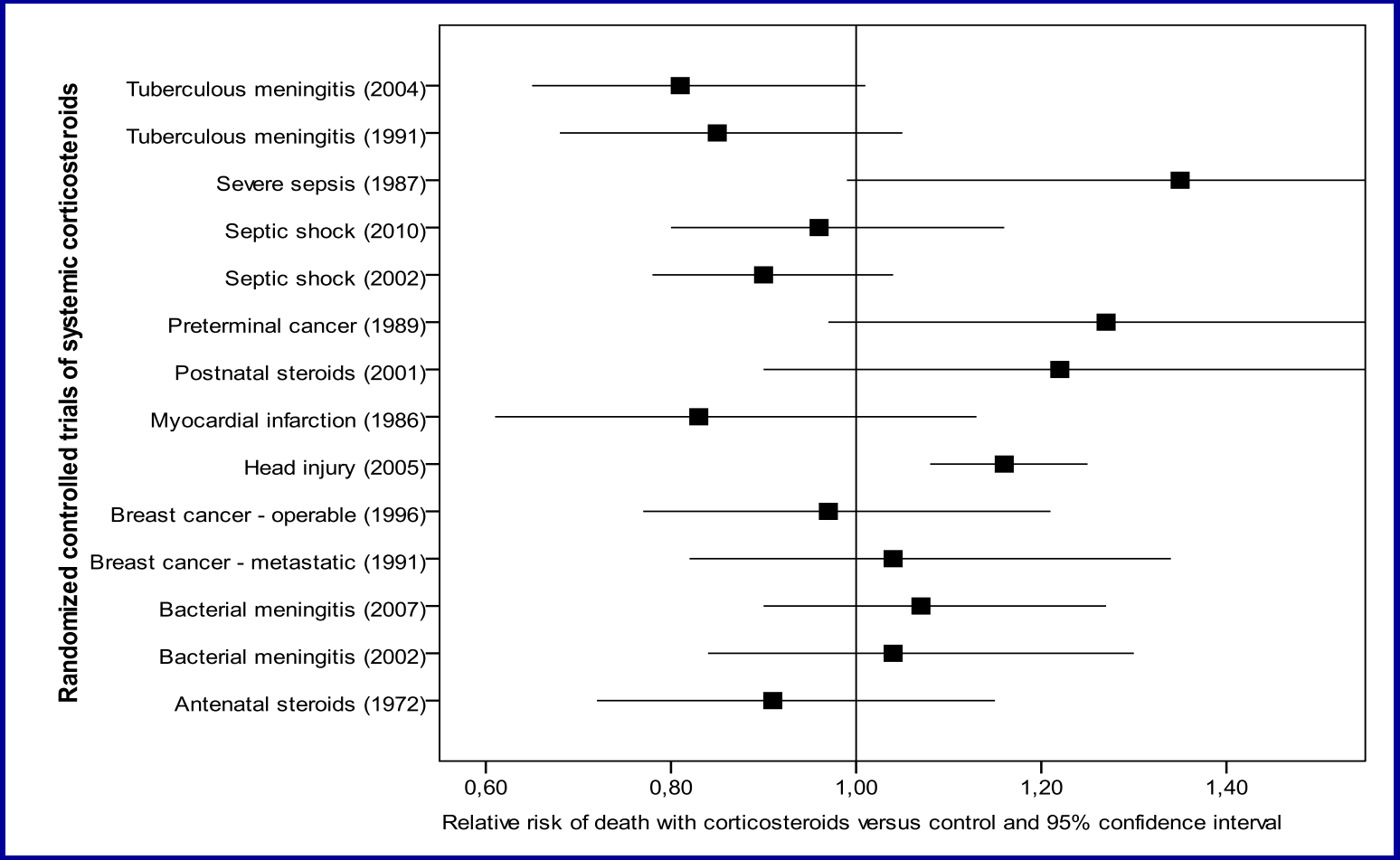


# 1200 (and counting) clinical trials of bevacizumab



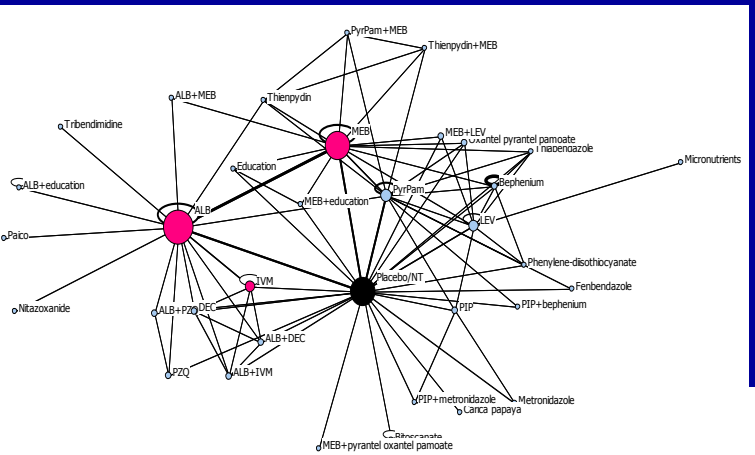

Meta-analysis forest plot for survival with bevacizumab *v* control in trials of patients with cancer. Each trial is shown by its year of publication, name of first author, and type of malignancy as well as the hazard ratio for survival and 95% confidence interval. Also shown are summary estimates including all trials and excluding the three trials stopped early, which showed large treatment benefits (Hurwitz 2004, Sandler 2006, Escudier 2007)

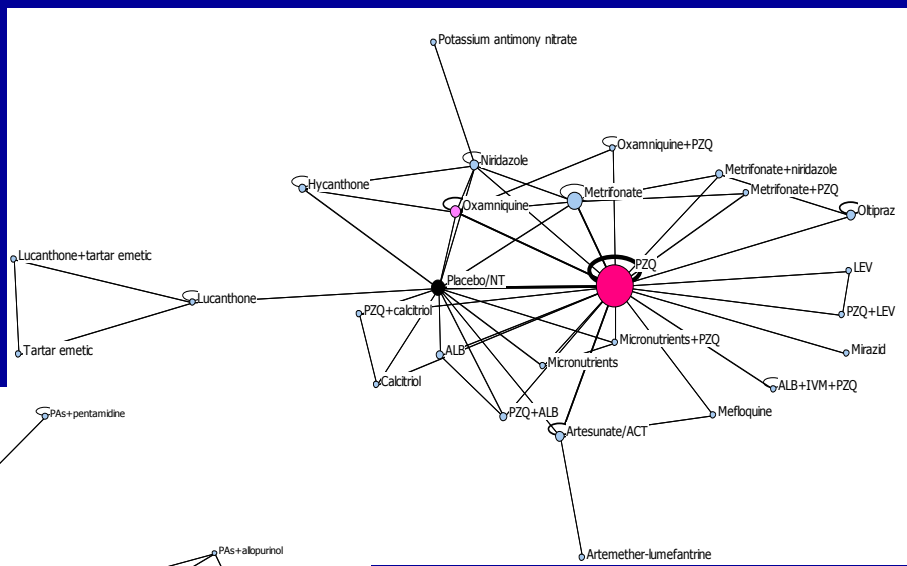
# Fifty years of research with 2,000 trials: 9 of the 14 largest RCTs on systemic steroids claim statistically significant mortality benefits

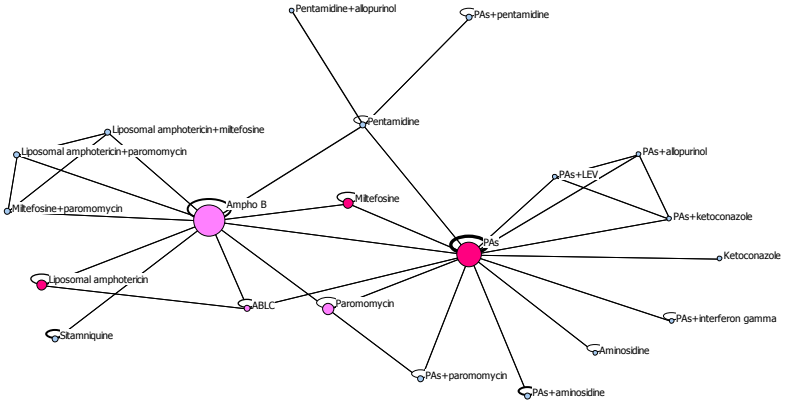


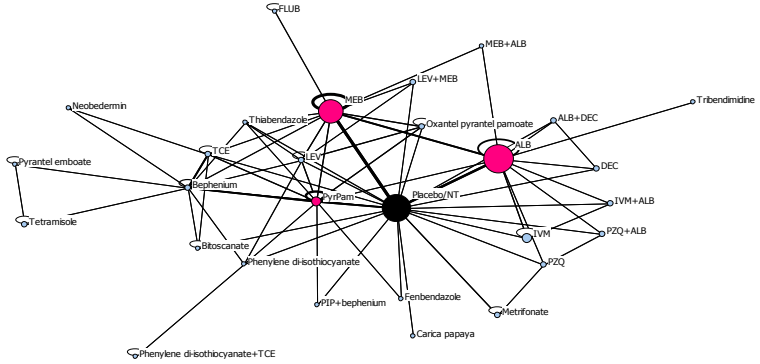

Contopoulos-Ioannidis and Ioannidis EJCI 2011

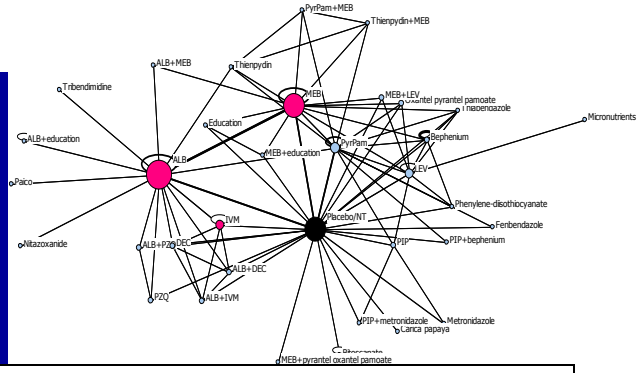
# Trial networks for neglected tropical diseases (burden: 1 billion people)




Kappagoda and Ioannidis, submitted

# What the next study should do?

- Maximize diversity
- Address comparisons that have not been addressed
- Minimize co-occurrence
- Break (unwarranted) homophily
- Be powered to find an effect or narrow the credible or predictive interval for a specific comparison of interest
- Maximize informativity across the network (entropy concept)
- Some/all of the above

# Maximizing entropy change in medical studies

The information gain (entropy change) from a new study is given by

$$DKL(p'||p) = w'\log(w'/w) + (1-w')\log((1-w')/(1-w)) + w'DKL(N(\mu',\sigma'^2)||N(\mu,\sigma^2))$$

The Kullback–Leibler divergence between the two normal distributions is given by

$$DKL(N(\mu',\sigma'^2)||N(\mu,\sigma^2)) = (\mu' - \mu)2 / 2\sigma^2 + ½ (\sigma'^2/\sigma^2 - 1 - \log(\sigma'^2/\sigma^2))$$

In case the major objective is to distinguish between a zero and a non-zero effect, the information gain of a result simplifies to

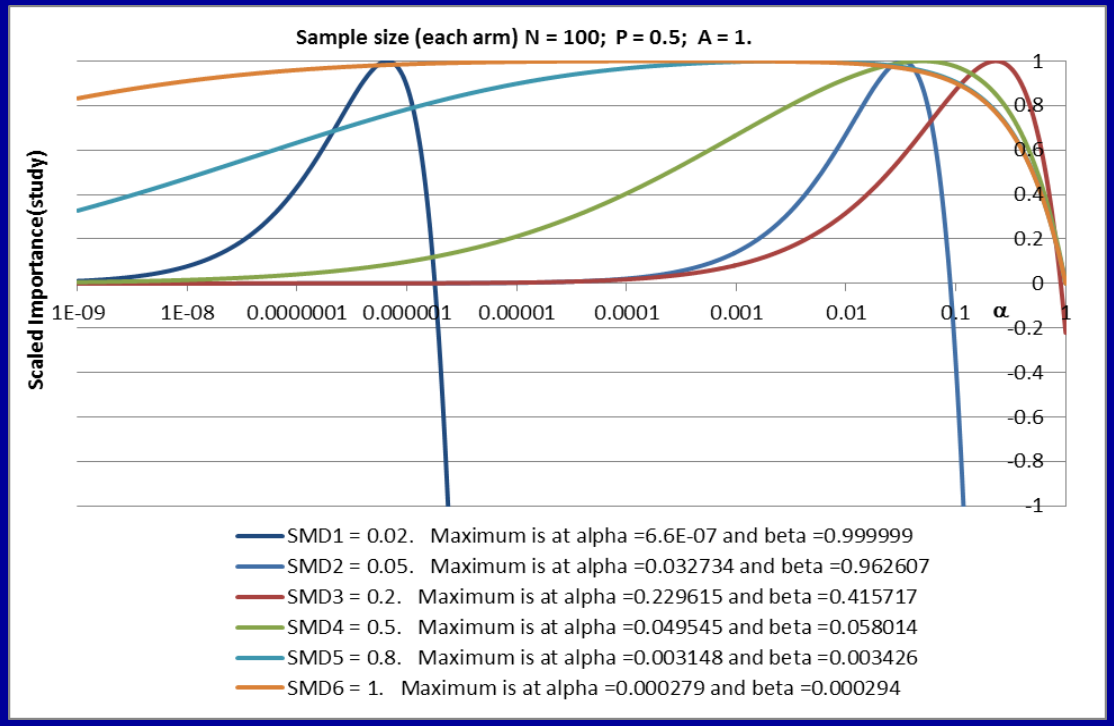
$$DKL(p'||p) = w'\log(w'/w) + (1-w')\log((1-w')/(1-w))$$

# Optimization function for the importance of a future study, taking into account the relative values of a TN, TP, FP, FN

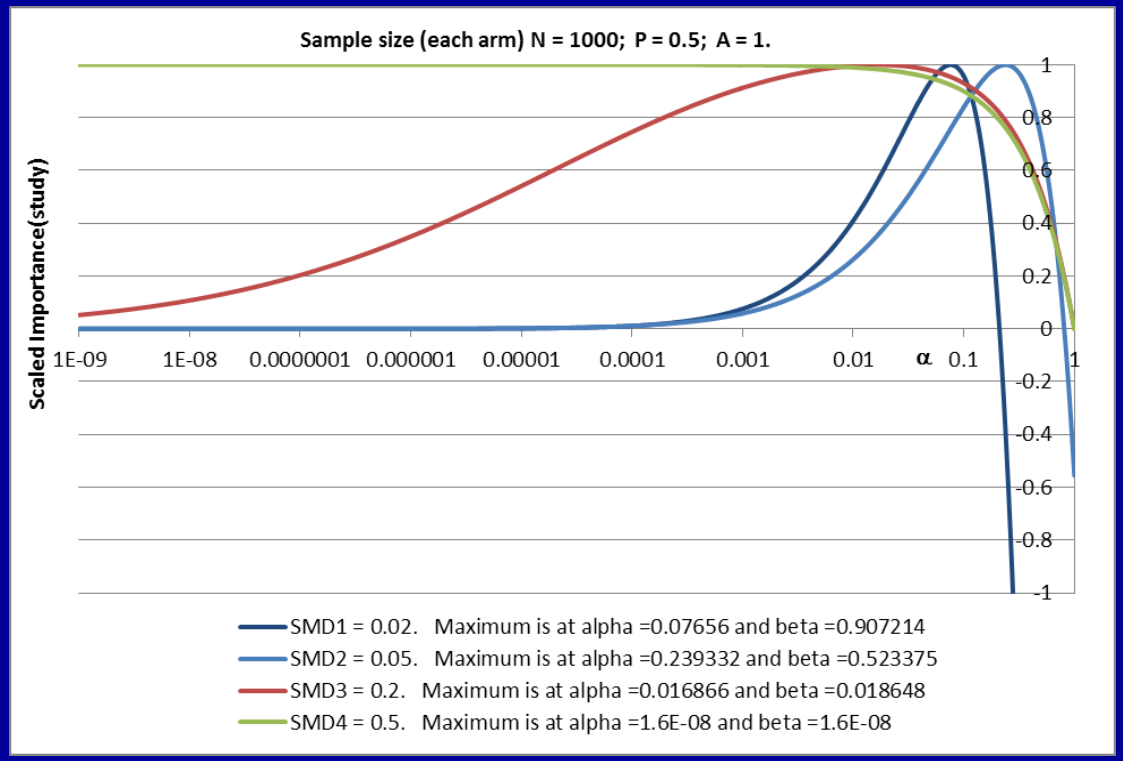
Some simple situations:

- Additive model with equal value assigned for TP, TN, FP, FN: $F(opt)= (-2\beta P-\alpha+\alpha P+P+1-P-\alpha+\alpha P)$
- Additive model with no value for true negatives: $F(opt)=P-2\beta P-\alpha+\alpha P$
- Additive model, at least one discovery is essential to make: $F(opt)=(P-2\beta P-\alpha+\alpha P)(1-\beta^{\Omega})$

# Additive optimization model for small randomized trial



# Additive optimization model for large randomized trial



# Meta-analysis=primary type of prospective research

We need to think about how to design prospectively large agendas of randomized trials and their respective networks for questions that are important to patients and can make a difference in their lives