Data sharing for clinical research: Experiences, techniques, methodologies

John Ioannidis

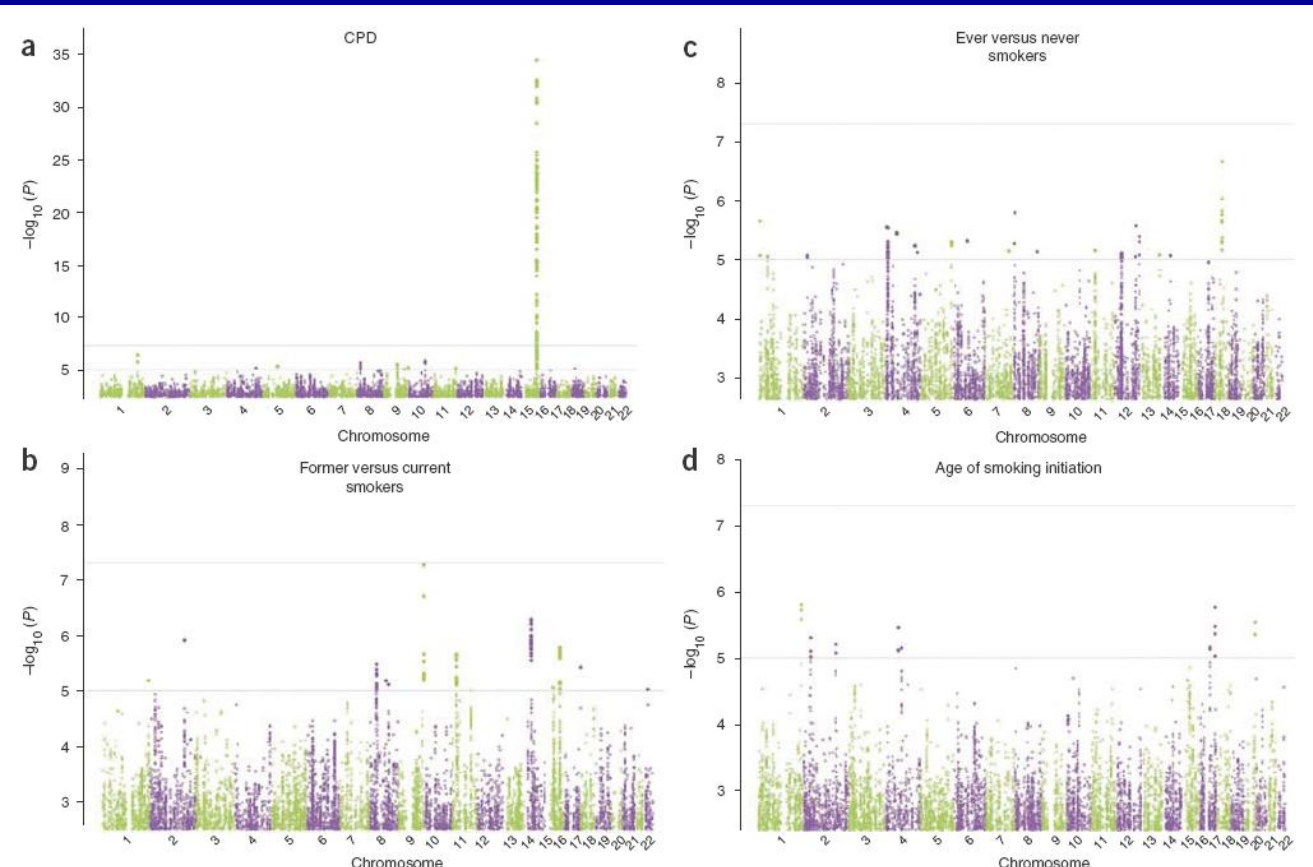# Several Libraries of Alexandria disappear every day

Some experiences in clinical research data sharing

- Meta-analyses of individual participant data from randomized trials for primary or secondary outcomes (efficacy or harms)

- Exploratory (discovery) analyses based on existing data, e.g. biomarkers

- Development and validation of predictive models

- Methodological work, e.g. risk modeling

- Repeatability and reproducibility checks

- Prospective meta-analyses and networks

Some basic principles

- It is more difficult (or impossible) to unearth data after the fact

- It is better to arrange upfront for the full individual-level data to be available

- Coding, cleaning, logical queries take some time, but their difficulty is probably over-rated

- Difficulty of standardization in post-hoc efforts limits the utility of the information

Prospective large-scale collaboration

nature genetics

Genome-wide meta-analyses identify multiple loci associated with smoking behavior
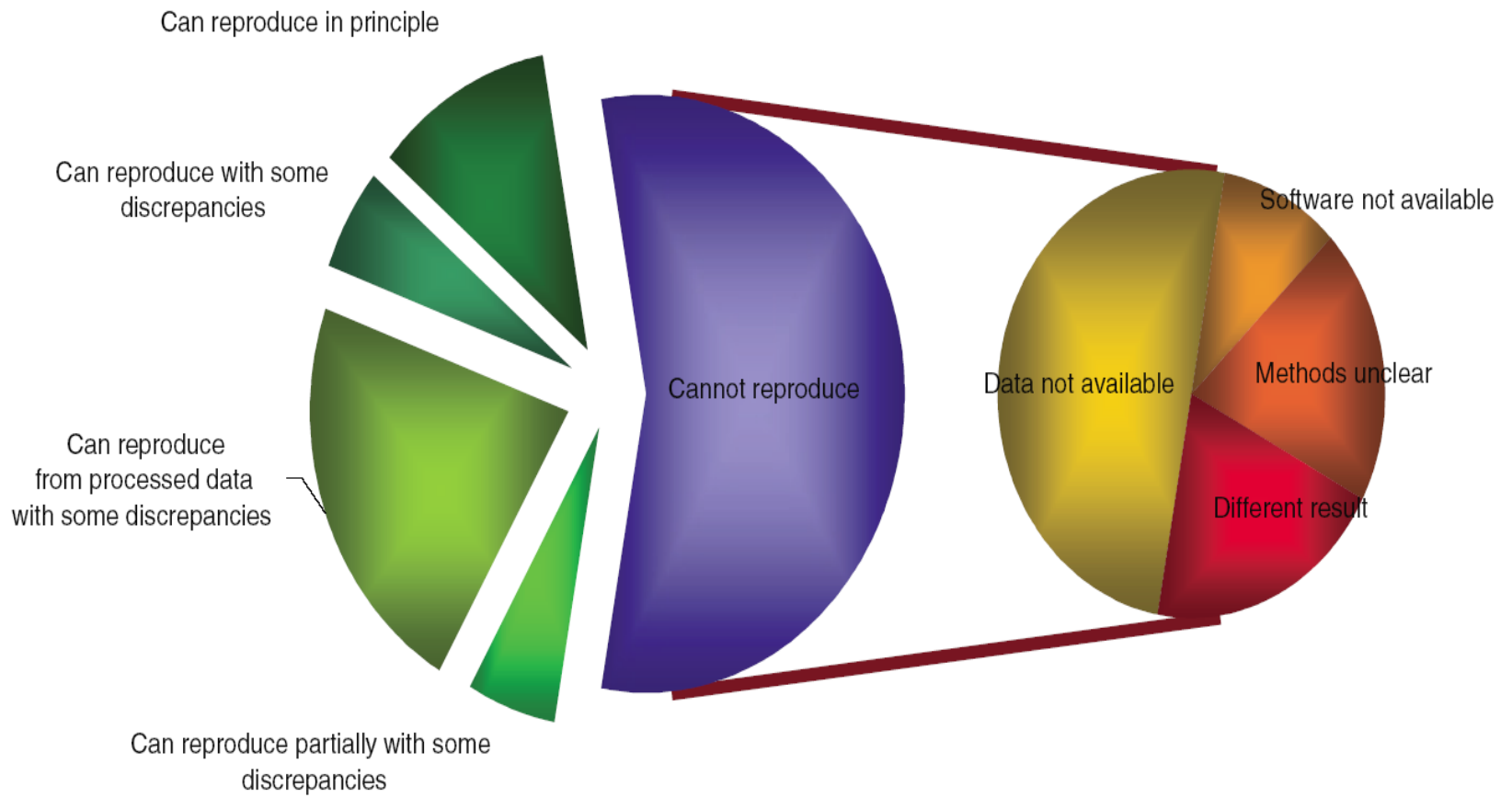
Nature Genetics, 2010

Public Availability of Published Research Data in High-Impact Journals

Alawi A. Alsheikh-Ali[1,2,9], Waqas Qureshi[3,9], Mouaz H. Al-Mallah[4,5], John P. A. Ioannidis[2,6,7,8,9]*

Journal	Impact Factor	Policy of Required Public Deposition for Types of Data				Policy of Provision of Materials and Methods			Full data deposited
		Microarray	Nucleic Acid	Protein	Macromolecular	Materials upon request	Protocols upon request	Condition of publication	Percentage of papers
New England Journal of Medicine	52.589								0
Cell	29.887								1
Nature	28.751								0
Lancet	28.638								0
Nature Medicine	26.382								0
Science	26.372								1
Nature Immunology	26.218								9
Nature Genetics	25.556								0
JAMA	25.547								1
Nature Biotechonology	22.848								5
Nature Materials	19.782								0
Immunity	19.266								0
Nature Cell Biology	17.623								0
Journal of Clinical Investigation	16.915								0
Archives of General Psychiatry	15.976								0
Journal of the National Cancer Institute	15.678								0
Nature Neuroscience	15.664								1
Journal of Experimental Medicine	15.612								0
Annals of Internal Medicine	15.516								0
Journal of Clinical Oncology	15.484								0
Nature Methods	15.478								6
Genes and Development	14.795								3
Nature Physics	14.677								0
PLoS Biology	13.501								2
Neuron	13.41								0
Molecular Cell	13.156								0
Circulation	12.755								0
PLoS Medicine	12.601								0
Developmental Cell	12.436								0
Gastroenterology	11.673								0
Genome Research	11.224								6
American Journal of Human Genetics	11.092								3
Nature Structural and Molecular Biology	11.085								0
Journal of the American College of Cardiology	11.054								0
Blood	10.896								0
Hepatology	10.734								0
Current Biology	10.539								0
Gut	10.015								0
British Medical Journal	9.723								0
Circulation Research	9.721								1
Plant Cell	9.653								0
Nano Letters	9.627								0
Journal of Cell Biology	9.598								0
PNAS	9.598								1
Molecular and Cellular Proteomics	9.425								7
PLoS Pathogens	9.336								0
American Journal of Psychiatry	9.127								0
American Journal of Respiratory and Critical Care Medicine	9.074								0
Annals of Neurology	8.813								0
PLoS Genetics	8.721								0

Repeatability

ANALYSIS

nature genetics

Repeatability of published microarray gene expression analyses

John P A Ioannidis[1–3], David B Allison[4], Catherine A Ball[5], Issa Coulibaly[4], Xiangqin Cui[4], Aedín C Culhane[6,7], Mario Falchi[8,9], Cesare Furlanello[10], Laurence Game[11], Giuseppe Jurman[10], Jon Mangion[11], Tapan Mehta[4], Michael Nitzberg[5], Grier P Page[4,12], Enrico Petretto[11,13] & Vera van Noort[14]
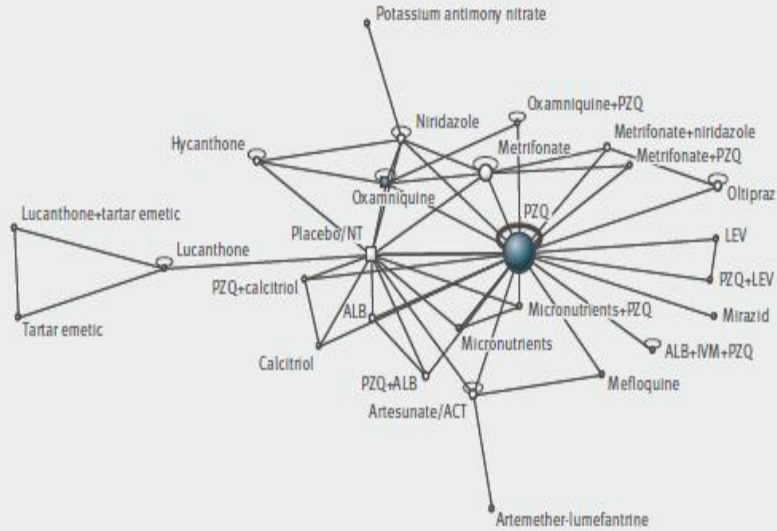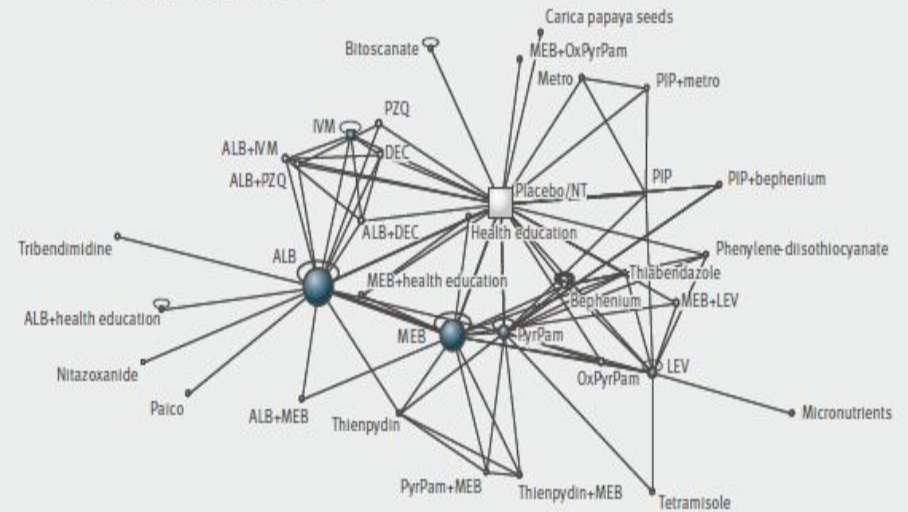
Figure 1 Summary of the efforts to replicate the published analyses.

Some additional problems encountered

- Politics

- Clash of evidence-based medicine with expert-based medicine (Goetzsche & Ioannidis, BMJ in press 2012)

- Not all datasets may be available at the individual participant level = new possibilities for selective reporting biases

- All of these problems are almost eliminated if sharing of individual level information is agreed before the data is generated
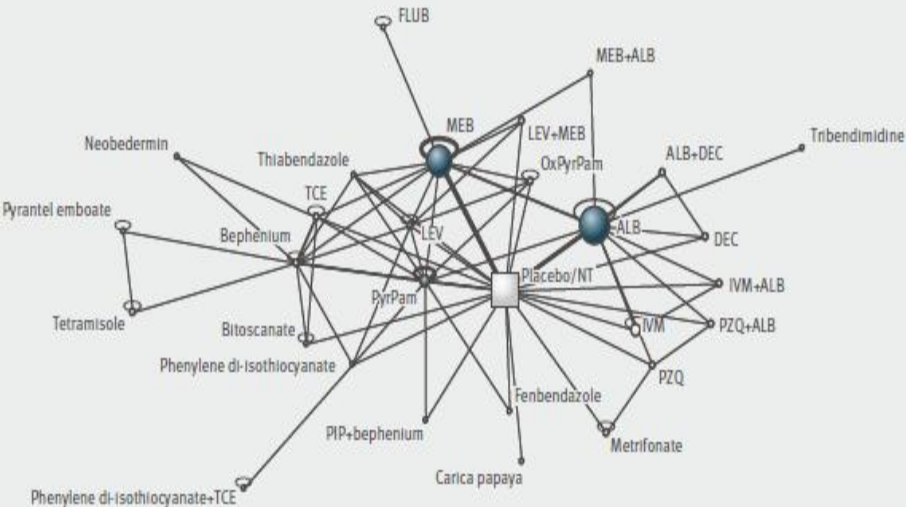
Schistosomiasis (113 trials)

Ascariasis (78 trials)

Hookworm (77 trials)

Trichuriasis (66 trials)

Kappagoda and Ioannidis, BMJ 2012 in press