Evidence and guidelines

Guidelines

"Systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances"

Institute of Medicine, 1990

Methods of Developing Guidelines

- **Informal consensus**
- **Formal consensus**
- **Evidence-based approach**
- **Hybrid approaches**

Consensus

Although it may capture collective knowledge, it is also vulnerable to the possibility of capturing collective ignorance

Grading quality of evidence and strength of recommendations (GRADE)

- How much confidence can we place in the recommendations?
- Clinical guidelines are only good as good as the evidence and judgments that are based on
- Systematic and explicit methods of making judgments can reduce errors and improve communication

BMJ 2004;328:1490-4

Formulation of guidelines: The need for research synthesis

- Health care decision makers need to access research evidence to make informed decisions on diagnosis, treatment and health care management for both individual patients and populations.
- There are few important questions in health care which can be informed by consulting the result of a single empirical study.

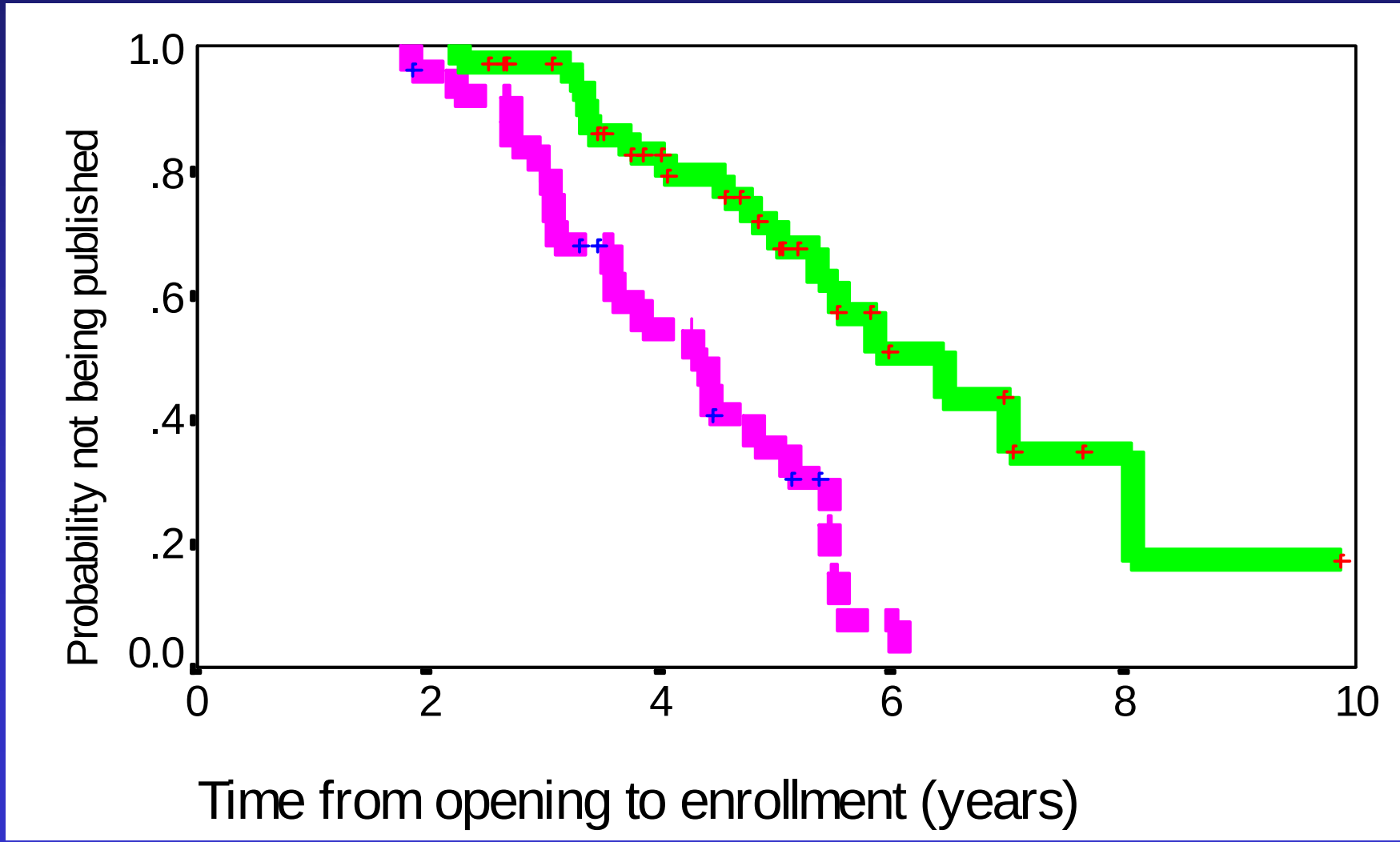
Field-wide issues in evidence synthesis

- Selection biases
- Early vs. late evidence
- Large vs. small studies
- Different study design effects
- “Quality” effects
- Heterogeneity and subgroups
- Overall validity of the research findings

Selection biases

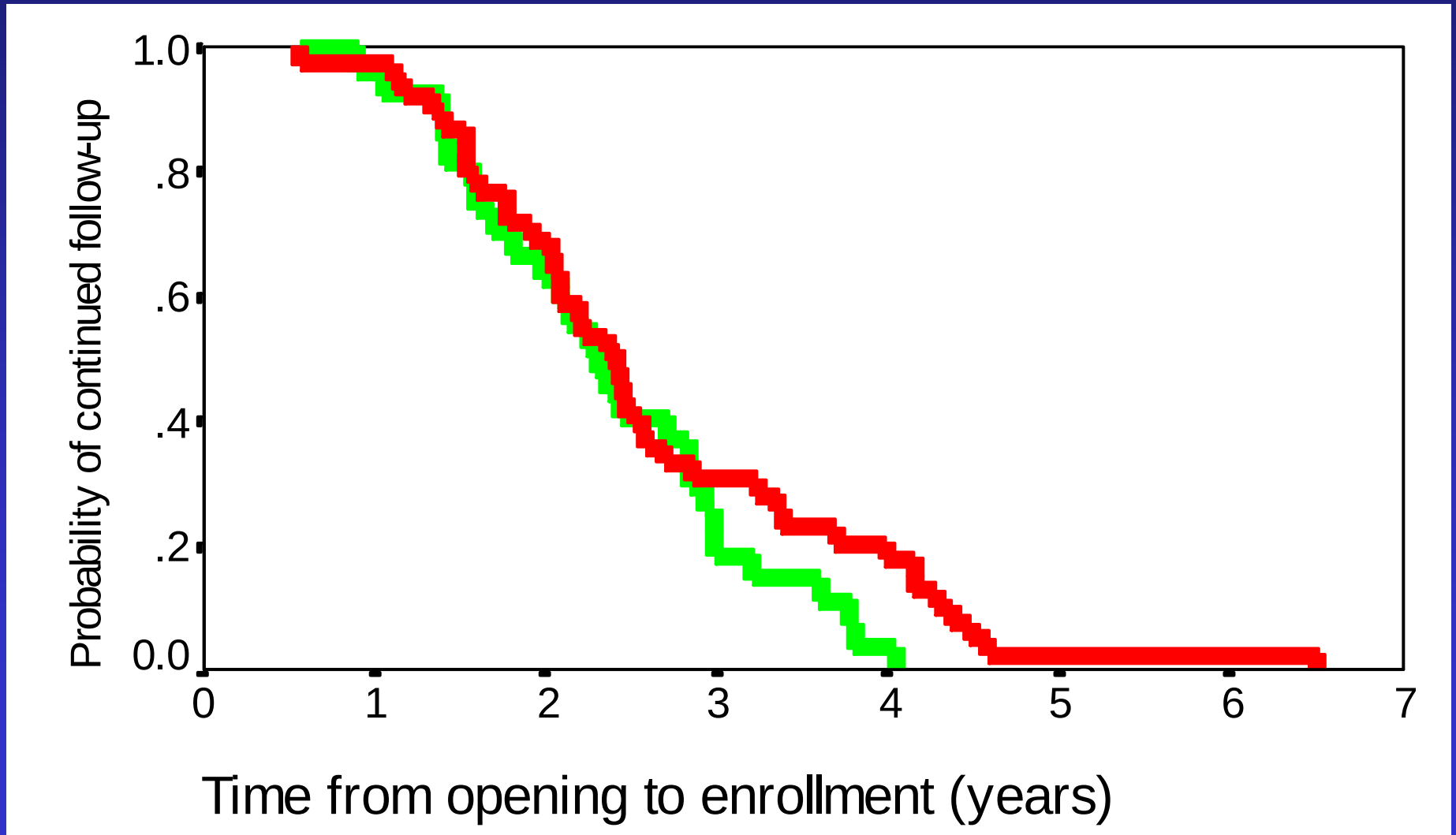
- Publication bias
- Time lag bias
- Selective outcome reporting bias

Time lag: bad news take longer to appear

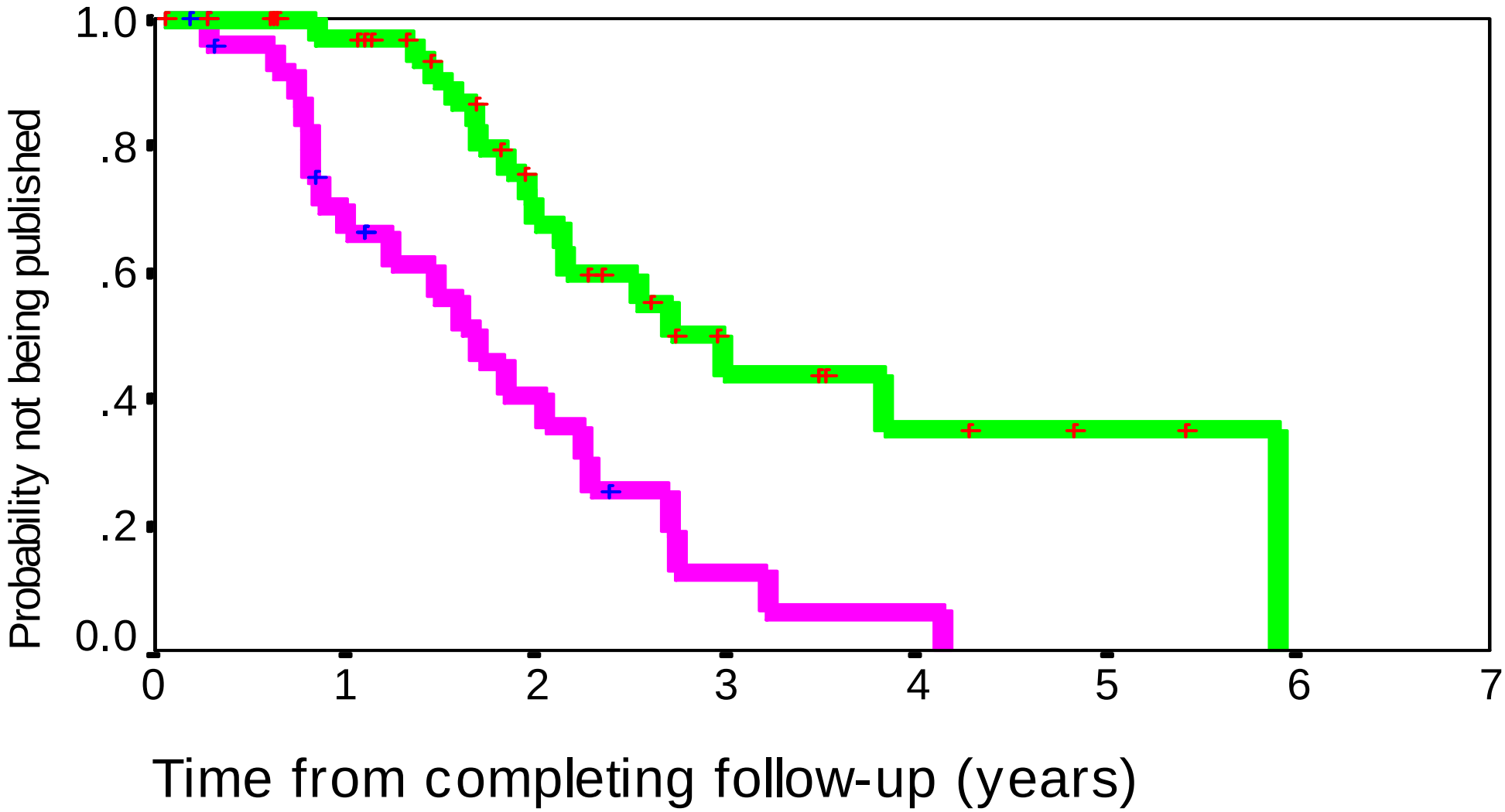


Ioannidis JP. JAMA 1998

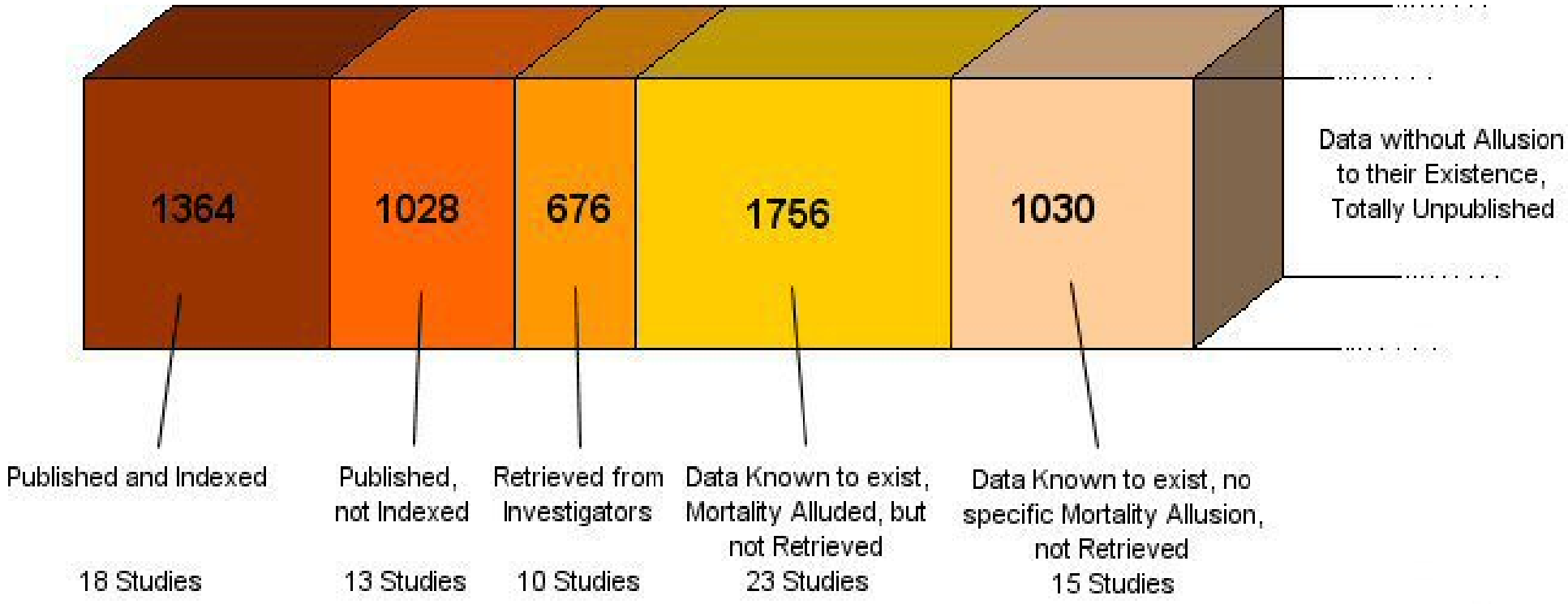
… even though they are obtained as fast..



…but publication is delayed



Prognostic factor meta-analysis: Readily available, available, hidden, and very well hidden data

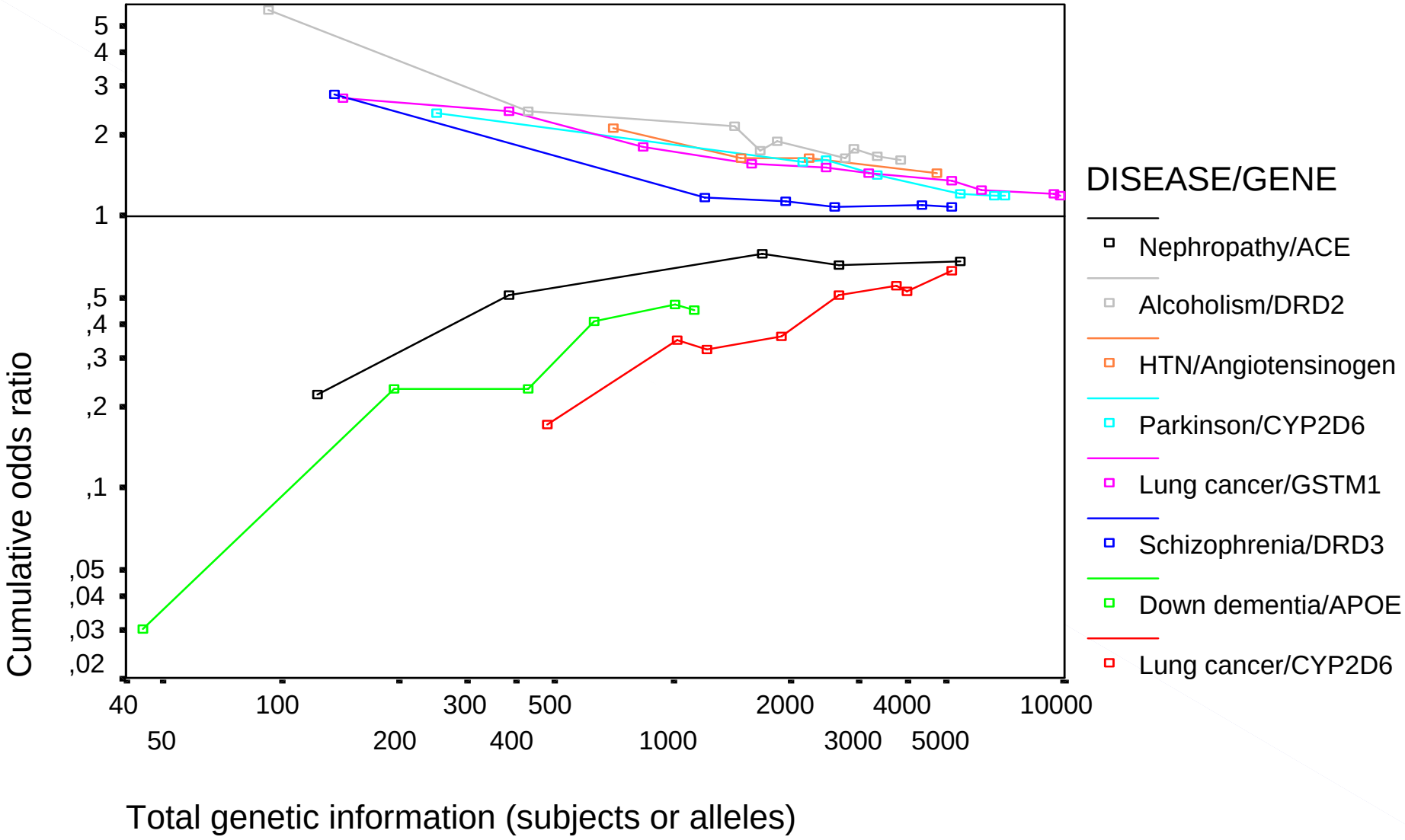


Kyzas, Loizou, Ioannidis, JNCI 2005

Early vs. late evidence

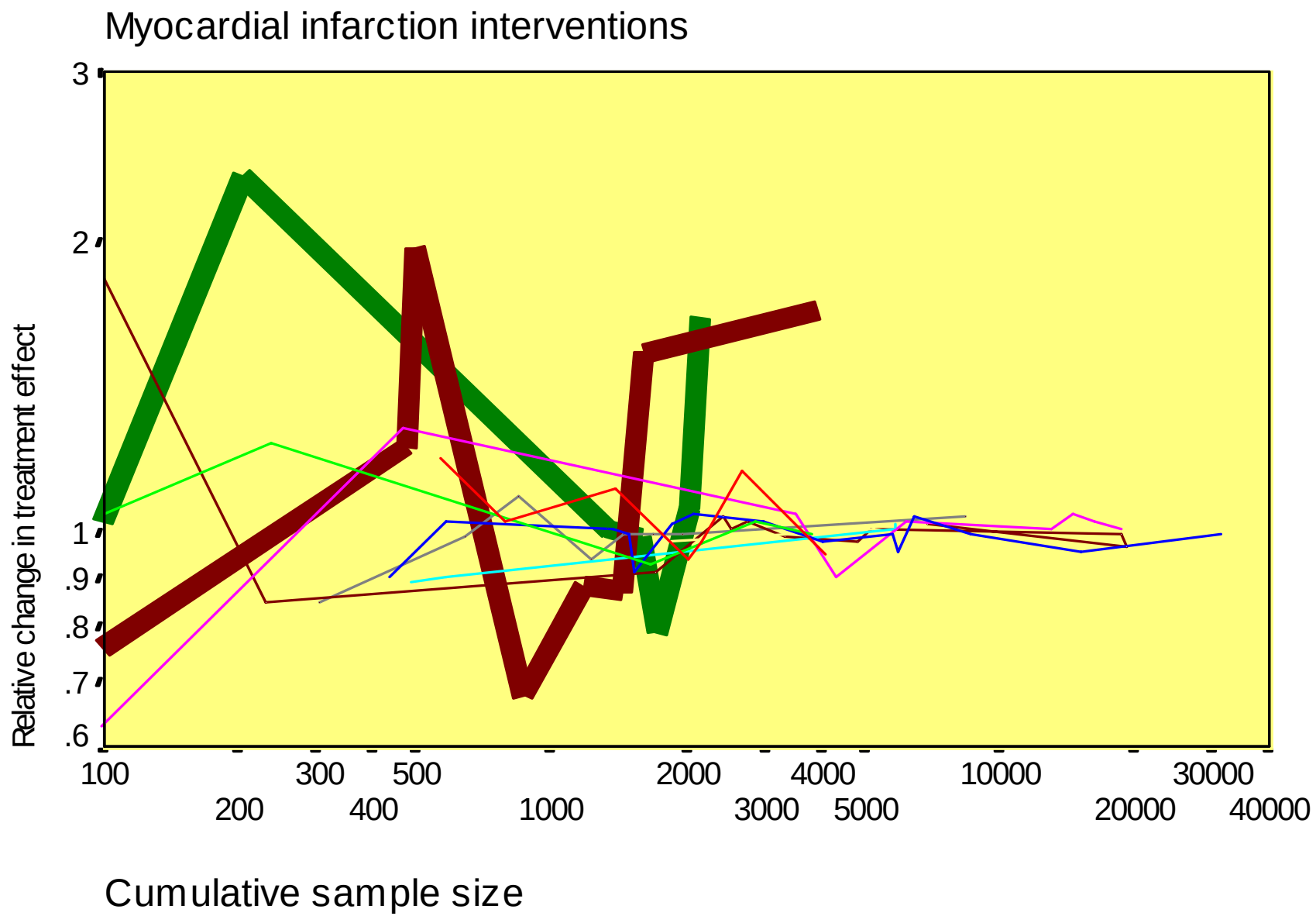
- Evidence evolves over time, it is never constant
- Evolution may change effect sizes
- Opposing effects may occasionally succeed each other in rapid sequence

Non-replicated diminishing effects



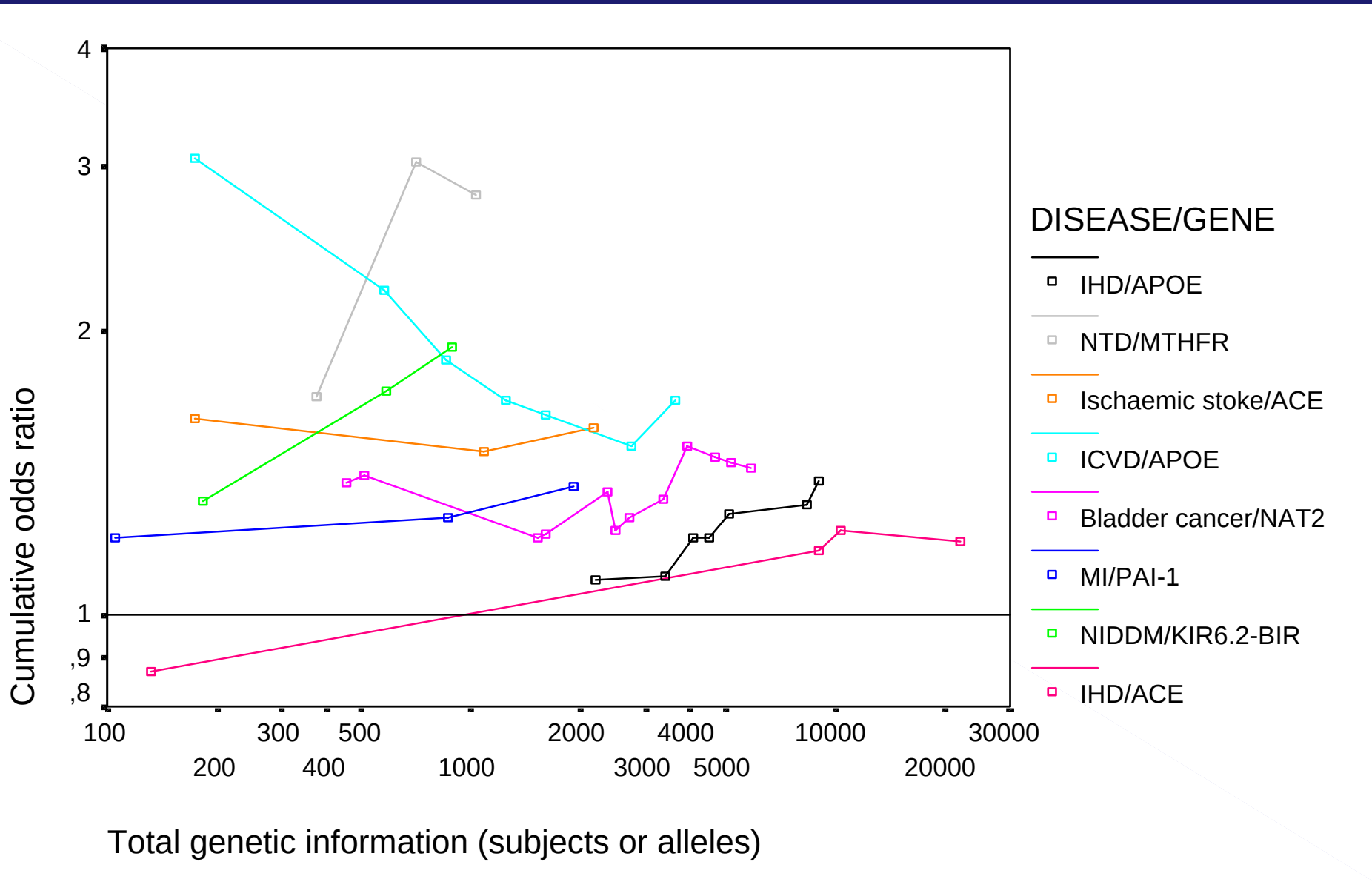
Ioannidis et al, Nature Genetics 2001

Discrepancies over time occur even in randomized trials



Ioannidis and Lau, PNAS 2001

The other side: don’t give up early



Large vs. small studies

- Theoretically they should not get different results
- Differences reflect both within study issues and field issues

Large vs. small studies in RCTs

- Empirical evidence shows that usually their results agree, but discrepancies may occur beyond change in 10-30% of the cases
- In these situations, large studies tend to give more conservative results, but this is not always the case
- Discrepancies tend to be more frequent for secondary than for primary endpoints

Ioannidis, Cappelleri and Lau, JAMA 1998

“Quality” of studies

- Some empirical evaluations have suggested that specific quality items such as lack of blinding and lack of allocation concealment in RCTs may inflate treatment effects (e.g. Shultz et al. JAMA 1995)
- It seems more likely that such quality deficits may be associated either with inflated or with deflated treatment effects (e.g. Balk et al. JAMA 2002)

Number of Systems to Rate the Quality of Individual Studies

Study Design (Grid)	Total	Scales, Checklists, and Component Evaluations	Guidance Documents	EPC Rating Systems
Systematic Reviews	20	11	9	0
Randomized Controlled Trials	49	32	7	10
Observational Studies	19	12	5	2
Diagnostic Tests	18	6	9	3

total =106

The two kinds of bad quality

- Quality is bad on (evil) purpose = the effect sizes are almost always inflated
- Quality is bad because of stupidity = the effect sizes may be anything; usually, but not always, they are deflated

Heterogeneity and subgroups

- Heterogeneity is very interesting: it may hint to both genuine diversity and bias
- Too much heterogeneity is suspect
- Too little heterogeneity may also be suspect
- Some heterogeneity is almost ubiquitous
- Over-interpretation through postulated subgroup differences can be dangerous

Overall validity

- Depends on the pre-evidence odds
- Depends on the evidence
- Depends on bias
- Depends on the field
- All of these may depend on each other

Contradiction in highly-cited clinical research on interventions

- Analyzed 115 articles published in 1990-2003 in the 3 major general medical journals (NEJM, JAMA, Lancet) and the top specialty journals that had received over 1000 citations each by august 2004
- Of those, 49 pertained to original assessments of interventions for therapy or prevention and 45 claimed effectiveness.
- Five of the 6 efficacy findings based on non-randomized trials were already contradicted or found to be exaggerated by 2004
- Even among highly-cited randomized trials, efficacy findings were already contradicted or found to be exaggerated in 9 of 39 interventions

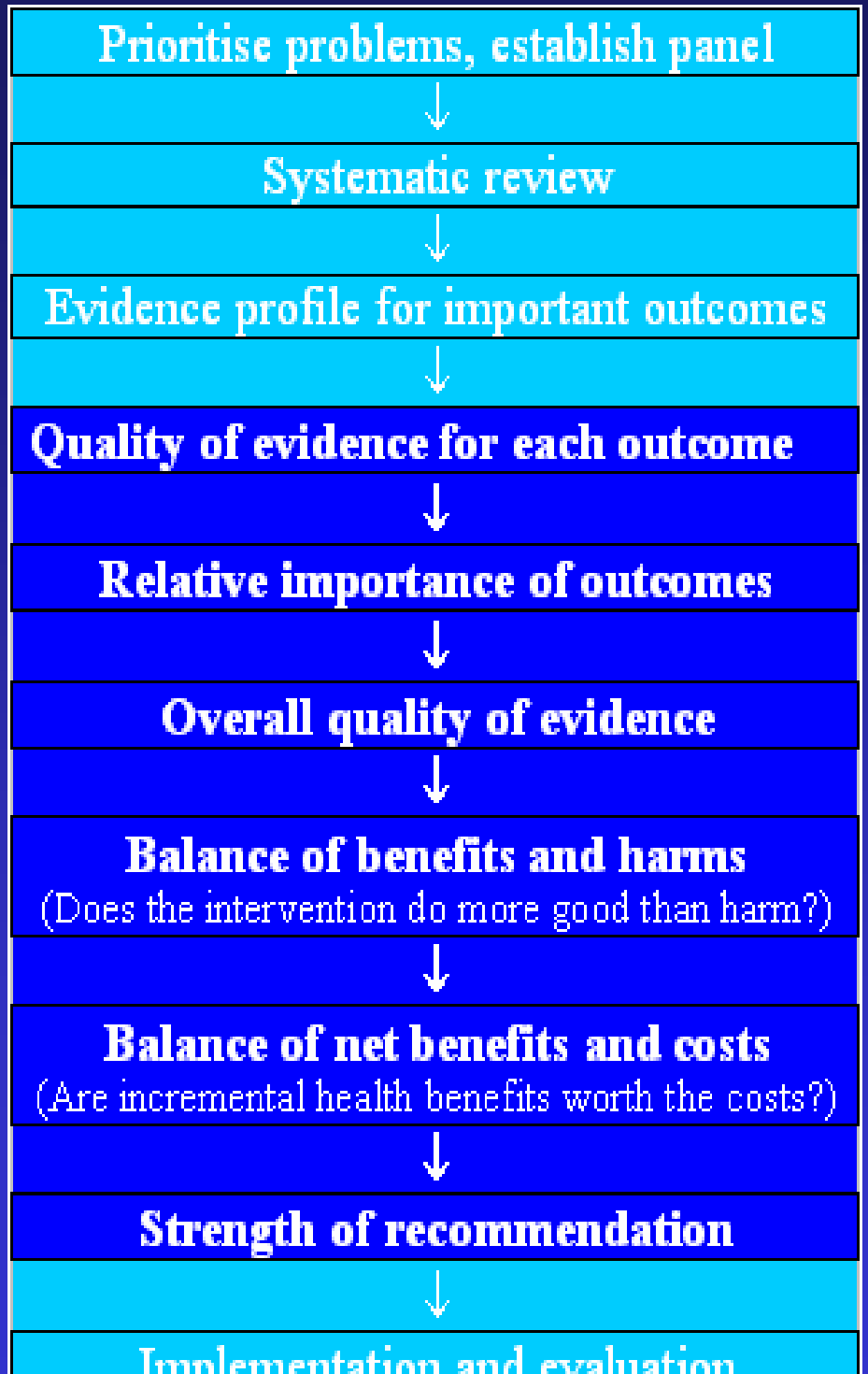
Ioannidis JP. JAMA 2005

Odds of a true finding are small

- When effect sizes are small
- When studies are small
- When field are “hot” (many teams work on them)
- When there is strong interest in the results
- When databases are large
- When analyses are more flexible

Ioannidis JP. PLoS Medicine 2005

Guidelines development proccess



After GRADE,2004

Quality assessment criteria

Quality of evidence	Study design	Lower if	Higher if
High	Randomised trial	**Study quality**: -1 Serious limitations -2 Very serious limitations -1 Important **inconsistency** **Directness**: -1 Some uncertainty -2 Major uncertainty -1 **Sparse or imprecise data** -1 High probability of **reporting bias**	**Strong association:** +1 Strong, no plausible confounders +2 Very strong, no major threats to validity +1 Evidence of a **Dose response gradient** +1 All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low	Any other evidence		

BMJ 2004;328:1490-4

GRADE definition of the categories of the quality of evidence

- High
  - Further research is very unlikely to change our confidence in the estimate of effect
- Moderate
  - Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- Low
  - Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- Very low
  - Any estimate of effect is very uncertain

GRADE Group

BMJ 2004;328:1490-4

APPRAISING CLINICAL PRACTICE GUIDELINES

AGREE instrument

- Download at www.agreecollaboration.com
- Original version and Greek translation
- Practice: apply the AGREE instrument to a guideline