Harms information: randomized evidence

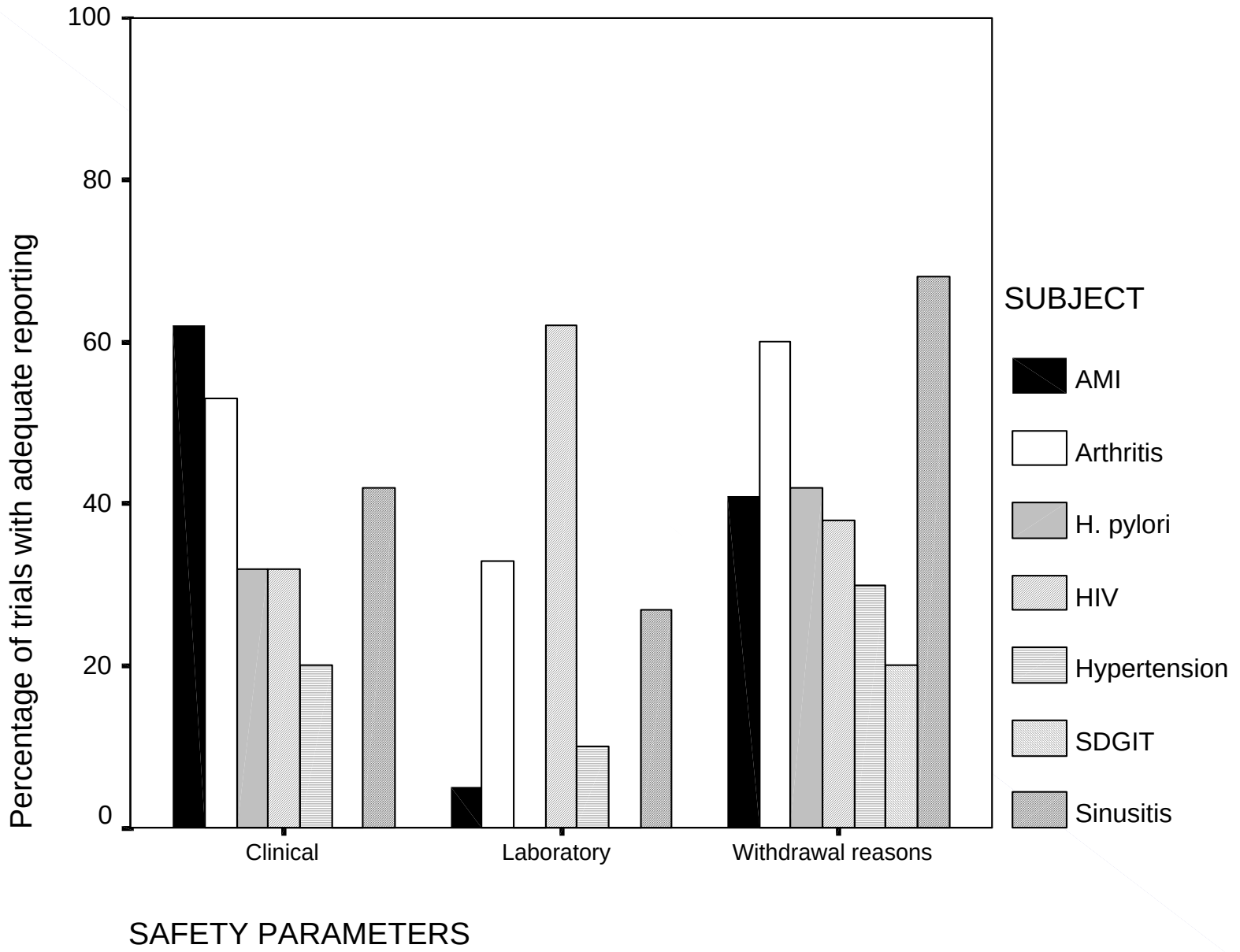
John P.A. Ioannidis

Harms

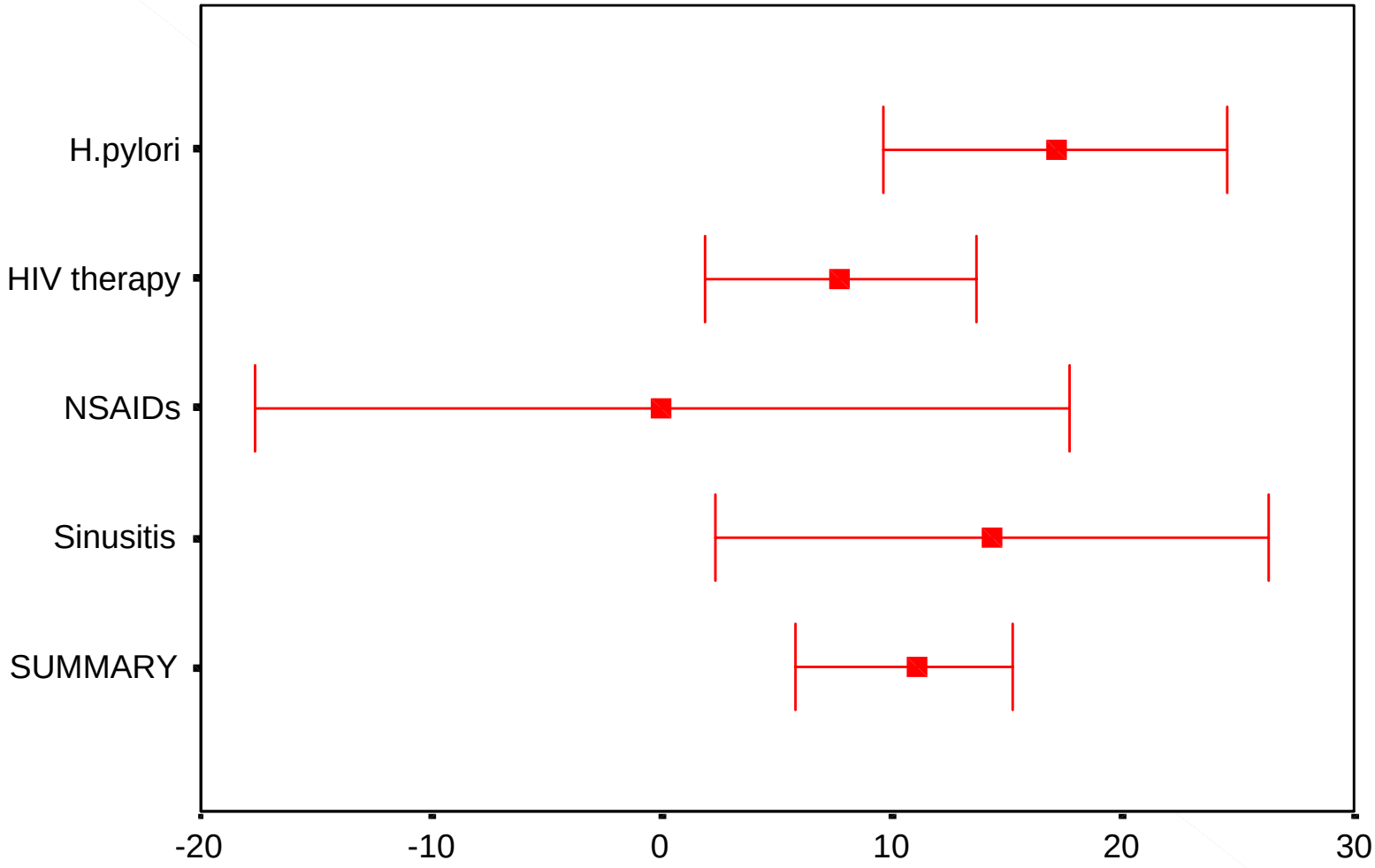
- Decision making in health care requires a balancing of efficacy information against safety information (=harms)
- Unfortunately there is increasing evidence that the recording of adequate information on harms in medical research, in particular randomized trials, is neglected

Harms reporting in RCTs

- In a study of 60 RCTs of drugs for the treatment of HIV infection (Ioannidis et al, Lancet 1998), the median space allocated to safety was only 0.5 pages per RCT report
- More than two-thirds of the RCTs did not mention adequately clinical side effects
- The space was less than the space allocated for the author names and their affiliations
- We replicated these findings across 7 medical fields and 192 trials (JAMA 2001)

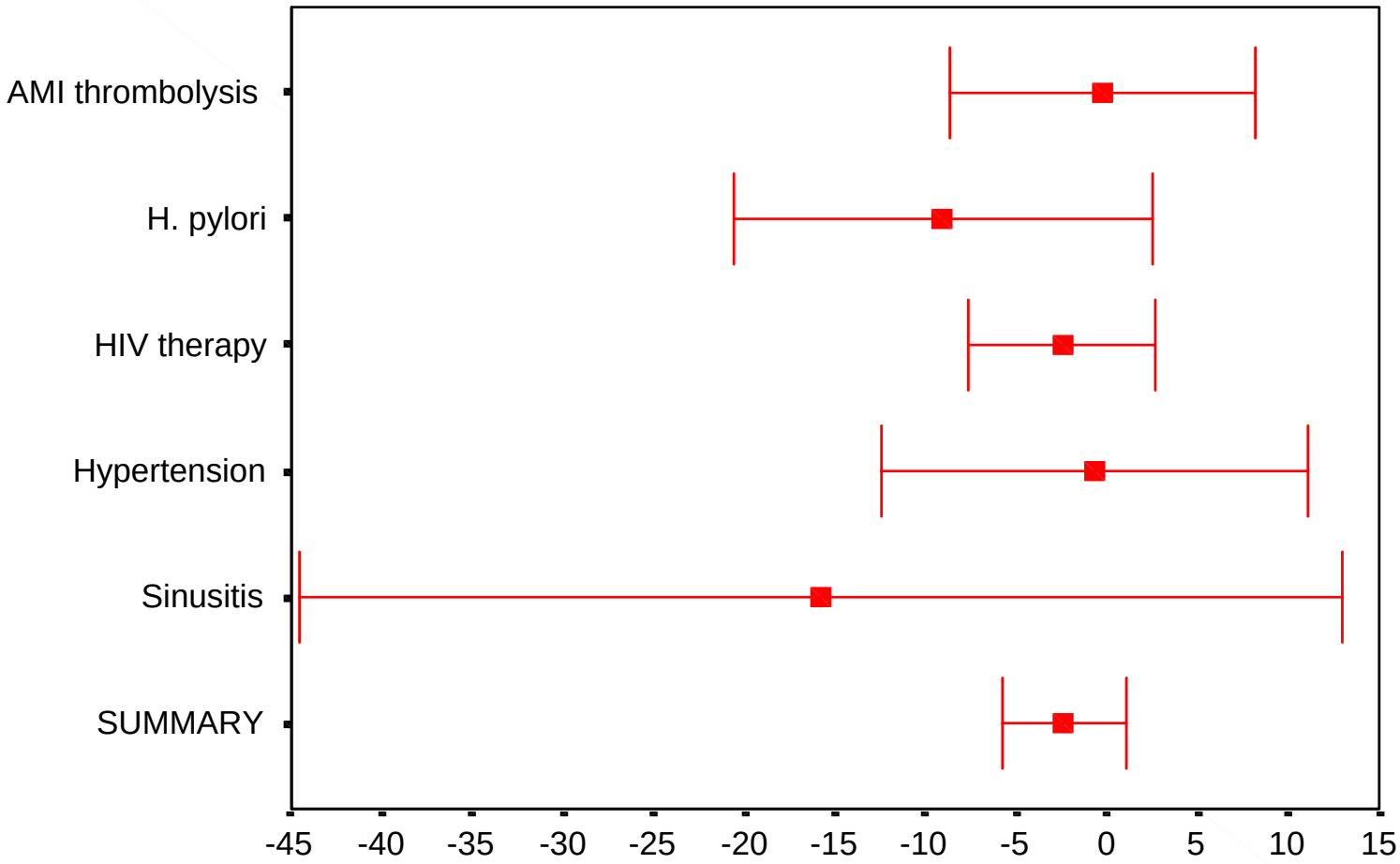

100
80
60
40
20
0
Percentage of trials with adequate reporting
Clinical
Laboratory
Withdrawal reasons
SAFETY PARAMETERS
SUBJECT
AMI
Arthritis
H. pylori
HIV
Hypertension
SDGIT
Sinusitis

DOSE COMPARISON INVOLVED

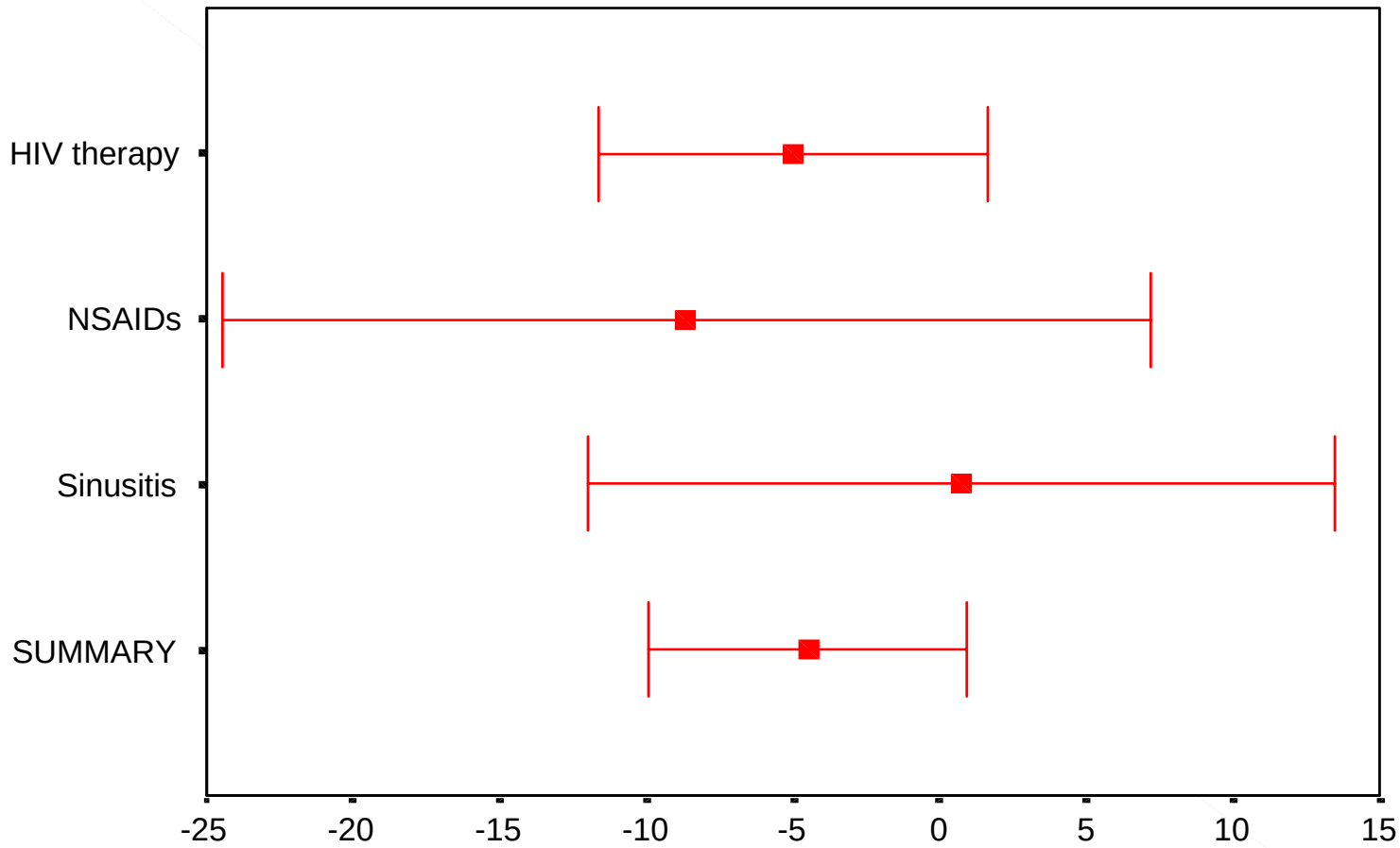


Increase in percentage of space devoted to safety

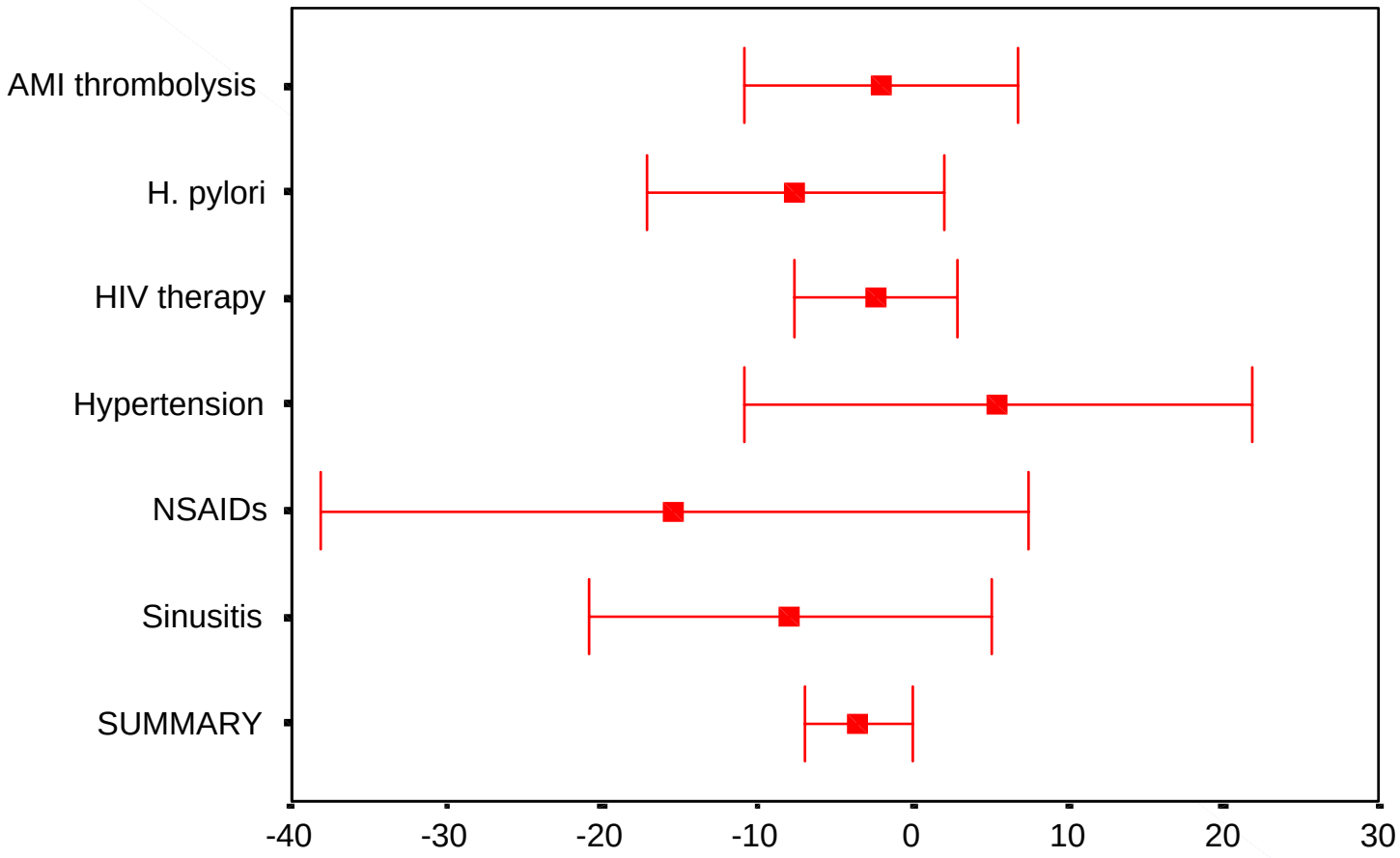
JOURNAL IMPACT FACTOR >7



PRIOR USE FOR OTHER INDICATION



SIGNIFICANT EFFICACY RESULTS



How feasible is meta-analysis of standardized safety data

- We investigated whether standardized safety data could at least be retrieved directly from trial investigators for the purpose of conducting meta-analysis
- We sent letters to investigators of all trials of antibiotics in acute sinusitis published in 1990-2000
- Acute sinusitis is a condition where clinical benefits of newer, expensive antibiotics are questionable or non-existent and knowledge of side effects is very important in decision making

Requested information

- Number of patients per trial arm with (a) nausea/vomiting, (b) diarrhea, and (c) both that resulted in hospitalization or at least use of intravenous fluids
- Number of subjects discontinuing study treatment for each of these three reasons per trial arm
- Cumulative days with each of these three toxicities in each arm

Trying to obtain safety information directly from trialists

- 16/38 trial investigators responded (42%), but only 9 of them provided some or all of the requested numerical information (although no one sent a protocol for safety data)
- Reasons for no data being contributed included: the information was not collected as part of the trial (n=2); the database had been lost (n=3); the database was difficult to locate, because it had been transferred to another company that had bought the drug (or even the whole company who had developed the drug originally) (n=2)
- Retrieved data included discontinuations and hospitalizations, but only 4 trialists could also give cumulative days of toxicity

Predictors: response / data retrieval

Predictor	Response	Data retrieval
	OR (p-value)	OR (p-value)
Time since original publication (per year)	0.82 (0.16)	0.75 (0.097)
Recorded sponsoring by non-industry source	2.31 (0.39)	6.75 (0.061)
European location of the first author	0.60 (0.45)	0.66 (0.59)
Sample size of the trial (per 100 patients)	1.13 (0.55)	1.35 (0.19)
Journal of publication (general med vs. specialty)	6.00 (0.047)	5.00 (0.061)

Sparse direct comparisons

- Despite a grand total of 7,434 randomized patients, more than 500 patients in direct comparisons were available only for the comparisons of loracarbef vs. doxycycline (1 trial, 662 patients) and cefixime vs. amoxicillin (4 trials with 652 patients)
- For the 1st comparison, the risk difference for clinical failures was 0.3% (-4 to 4.7) and for discontinuations due to toxicity 0.3% (-2.3 to 2.8)
- For the 2nd comparison, clinical failures decreased by 0.7% (-4.8 to 5.6) with cefixime and toxicity discontinuations increased by 1.2% (-2.6 to 5.0)

Other sources of heterogeneity for harms-related data

- Use of concomitant medications: in all patients; allowed; discouraged but recorded; not mentioned
- Mode of collection of safety information: passive; active with indirect questioning; active with direct questions; not mentioned
- Variable duration of treatment and variable period of surveillance for adverse events (3-35 days; usually not mentioned)
- Use of blinding, especially patient blinding
- Selection criteria for the study population
- Intention-to-treat vs. on-treatment approach

Harms are neglected

- Availability of standardized harms data for meta-analyses is limited even for recent trials and even when special efforts are made to retrieve them
- Harms-related data are poorly collected, haphazardly reported and easily forgotten and buried.
- The composite evidence is blurred by the fragmentation of trial comparisons testing several agents and using spurious "equivalence" efficacy designs, while neglecting safety
- For uncomplicated acute sinusitis at least, the risks of new antibiotics may equal or outweight their questionable benefits

CONSORT

- Original statement published in JAMA in 1996
- Revision published concomitantly in several medical journals in 2001
- Widely accepted as the gold standard for reporting of RCTs across hundreds of medical journals

CONSORT 1996

- No mention of harms

CONSORT 2001

- Item 19 added: “All important adverse events or side effects in each intervention group.”

Neglecting harms

- Among 375,143 entries in the Cochrane Central Register of Controlled Trials (Cochrane Library, issue 3, 2003), the search terms *harm* OR *harms* yielded 337 references
- This is compared with 55,374 retrieved using *efficacy* and 23,415 retrieved using *safety*.
- Of the 337, excluding several cases articles on self-harm or harm-reduction (an efficacy-equivalent term), there were only 3 trial reports and two abstracts that had these words in their titles.
- Of the 3 trial reports, one started with the clause "*more good than harm*" and the other two actually focused on the harms of the placebo arm

Words can be important

- **Harms** represent the totality of possible adverse consequences of an intervention or therapy; they are the direct opposite of **benefits** against which they must be compared.
- **Safety** implies substantive evidence of an absence of harm; the term is often misused when there is simply absence of evidence of harm.
- **Side effects** describe unintended drug effects. The term is not successful, however, at conveying harm, as some side effects may be beneficial. Furthermore, it tends again to understate the importance of harms, because *side* may be perceived as denoting secondary importance.

CONSORT for harms

- Initial meeting Barcelona, September 2001
- Working period 2001-2003
- Final consensus meeting, Montebello, May 2003
- Published in Annals of Internal Medicine November 2004

To avoid - 1

- Using generic/vague statements such as “the drug was generally well tolerated” or “the comparator drug was relatively poorly tolerated”
- failing to provide separate data for each arm
- providing summed up numbers for all adverse events per arm, without separate data for each type of adverse event
- providing summed up numbers for a specific type of adverse event regardless of severity or seriousness
- reporting only the adverse events that have been observed in a certain frequency or rate threshold (e.g. over 3% or 10% of participants)
- reporting only the adverse events that reach a p-value threshold in the comparison of the randomized arms (e.g. $p<0.05$)

To avoid - 2

- reporting measures of central tendency (e.g. means or medians) for continuous variables without any information on extreme values
- handling improperly or disregarding the relative timing of the events, when timing is an important determinant of the adverse event in question
- not distinguishing between patients with one and patients with multiple adverse events
- providing statements about the presence or not of statistical significance without giving the exact counts of events
- not providing data on harms for all randomized patients

Large-scale evidence on harms from meta-analyses

- We aimed to assess how frequently systematic reviews of randomized controlled trials convey large-scale evidence on specific, well-defined adverse events.
- To do this, we screened the entire Cochrane Database of Systematic Reviews for reviews with quantitative data on specific, well-defined harms for at least 4000 randomized subjects, the minimum sample required for adequate power to detect an adverse event occurring in 1% with an intervention.
- Main outcome measures included the number of reviews with eligible large-scale data on harms, number of reviews being ineligible for various reasons, and magnitude of recorded harms (absolute risk, relative risk) based on large-scale evidence.

Large-scale evidence on harms from meta-analyses

- We aimed to assess how frequently systematic reviews of randomized controlled trials convey large-scale evidence on specific, well-defined adverse events.
- We screened the entire Cochrane Database of Systematic Reviews for reviews with quantitative data on specific, well-defined harms for at least 4000 randomized subjects (power 80%, alpha=0.05, rate difference 1%, control rate <0.25%)
- Main outcome measures:

  number of reviews with eligible data on harms
  number of reviews being ineligible- and reasons

  magnitude of recorded harms on large-scale evidence

Large-scale evidence on harms

- Across 1727 Cochrane reviews, 138 included randomized evidence on ≥4000 subjects.
- Only 25 (18%) of them had eligible harms data, as defined above, while 77 had no data on adverse events and 36 had data on harms that were either non-specific and/or pertained to <4000 subjects.

CHARACTERISTIC	N (%)
Type of disease or condition	
Infectious diseases*	34 (24.6)
Perinatal conditions and pregnancy	29 (21)
Cancer	16 (11.6)
Cardiovascular diseases	11 (8)
Cerebrovascular diseases	10 (7.2)
Smoking	8 (5.8)
Bone and joint diseases	5 (3.6)
Mental health	4 (2.9)
Other	21 (15.2)
Interventions	
Vaccines or passive immunization	10 (7.2)
Anti-infective agents	20 (14.5)
Antiplatelet agents or anticoagulants	10 (7.2)
Other drugs	33 (23.9)
Surgical or other invasive procedures	8 (5.8)
Non-invasive technology	13 (9.4)
Behavioral/psychotherapy/counseling	14 (10.1)

Why is large-scale evidence on harms not available

REASON FOR LACK OF ELIGIBILITY	N (%)
No separate quantitative data on any adverse event	77 (68.1)
Quantitative data on ?4000 subjects only on non-specific adverse events	17 (15.0)
Composite counts of several different adverse events	8 (7.1)
Lack of grading	2 (1.8)
Reporting only aggregate withdrawals due to toxicity	4 (3.5)
Combination of reasons above	3 (2.6)
Quantitative data on specific adverse events on <4000 subjects	10 (8.9)
Quantitative data on non-specific adverse events on <4000 subjects	9 (8.0)

Large-scale evidence: Statistically significant differences for harms

- Of 66 specific harms with adequate data addressed in the 25 eligible reviews, 25 showed formally statistically significant differences between compared arms
- Most pertained to serious or severe adverse events and absolute risk differences <4%

Examples

- Convulsions with whole cell vs. acellular pertussis vaccines:
  RD 0.04% (0.01-0.07), RR 2.1 (1.4-3.2)
  15 trials with 124387 subjects
- Major extracranial bleed with anticoagulants in presumed acute ischemic stroke:
  RD 0.91% (0.67-1.16), RR 3.3 (2.4-4.7)
  15 trials with 22794 subjects
- Severe skin rash with clopidogrel vs. aspirin to prevent vascular events:
  RD 0.16% (0.04-0.28), RR 2.5 (1.2-5.2)
  1 trial of 19185 subjects
- Hepatitis with isoniazid for TB prophylaxis in HIV-negative people:
  RD 0.45% (0.31-0.60), RR 5.5 (2.6-12)
  1 trial with 20874 subjects

Information from large trials not reaching systematic reviews

- We screened the 113 systematic reviews that did not present specific large-scale evidence on harms for the largest randomized trial included in each of them.
- We identified 31 trials with a sample size of at least 4000 randomized subjects that had been published in a journal with impact factor >1.
- Among these 31 trials, nine (29%, 95% CI, 14-48%) presented detailed enough data on specific harms that would qualify for our definition of large-scale evidence on specific harms, but had nonetheless not been included in the systematic reviews.

Information from large trials not reaching systematic reviews

- Available harms data were not conveyed from the randomized trial to the systematic review (n=5 cases):
  - complications of heptavalent pneumococcal vaccine
  - digoxin toxicity
  - toxicity of calcium and aspirin for pre-eclampsia px
  - adverse reactions to antihypertensive agents
- Conveyed information lost its specific, well-defined quality (n=4 cases):
  - necrotizing enterocolitis with antibiotics for preterm labour
  - adverse events of cholera vaccine
  - salmeterol vs. salbutamol in asthma

Concluding comments

- Unfortunately, information on harms is poorly reported in randomized trials and further wasted in systematic reviews and meta-analyses thereof
- The situation may be improved at the design and reporting level, and also at the level of systematic reviews
- Systematic meta-analyses may provide reliable large-scale randomized evidence on well-defined harms