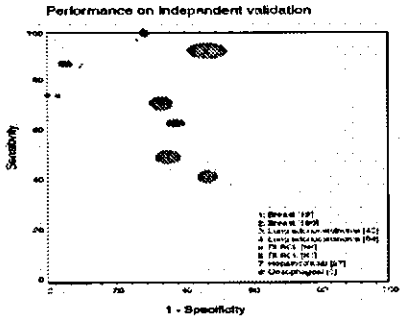
A survey of 37 meta-analyses of prognostic factors for cancer survival

- 28 meta-analyses (76%) were eventually found to be statistically significantly related with mortality.
- Only 2 meta-analyses (13%) stated an effort to retrieve data from the primary investigators, and only 1 of them actually presented the number of patients for whom data were retrieved.
- Another 9 meta-analyses reported on the number of studies with eligible but not evaluable data; however this was limited to studies that reported survival data in non-usable form and did not consider studies that clearly had collected follow up information, but presented no survival data at all in their publications.
- Only 5 meta-analyses (28%) used a standardized follow-up time, and only 1 (6%) tried to use a standardized definition for the expression of the prognostic factor, to the extent that this was possible. None however managed to convert the data from all studies to exactly the same definition.
- Five papers (28%) considered publication bias; in three of them the applied test was statistically significant, while in another one it was falsely claimed that there was no evidence of funnel plot asymmetry.

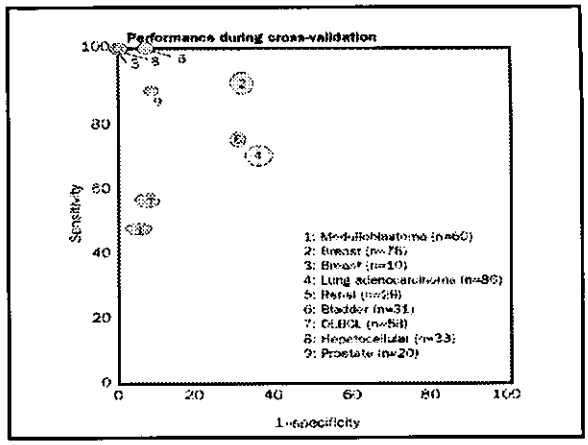
Microarrays and large-scale discovery research

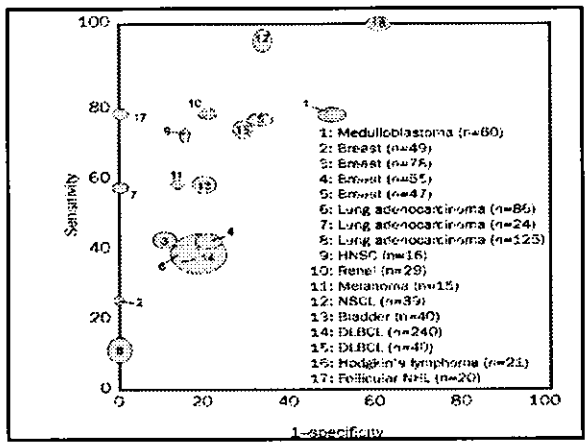
- "The Biological Periodic Table will not be two-dimensional, but will reflect similarities at diverse levels. It will be necessary to take global views of biological processes: simultaneous readouts of all components" (Lander E. Array of hope. Nat Genet 1999)
- "The purpose of this work is to find treatments or cures for all human diseases by the year 2050" (Schena M, Microarray analysis, 2003)

Small promising studies: microarrays



Ntzani and Ioannidis, Lancet 2003





Predictors of significant associations between molecular subtypes and outcomes

CHARACTERISTIC	ODDS RATIO (95% CI)	P-VALUE
Malignancy type (hematologic vs. solid tumor)	1·89 (0·65-5·47)	0·24
Sample size for specific cancer (2-fold increase)	3·39 (1·63-7·08)	0·001
Microarray type (oligonucleotide vs. cDNA)	1·61 (0·67-3·88)	0·29
Microarray probes (10-fold increase)	6·22 (1·89-20)	0·003
Probe type (fluorescent vs. radioactive)	0·17 (0·04-0·83)	0·03
Not reported separation	1·14 (0·43-3·02)	0·79
Amplification apparently used	0.99 (0.54-1.83)	0.98
Laser capture microdissection used	0·59 (0·05-6·80)	0·67
Paired normal tissue control	0·78 (0·26-2·33)	0·65
Number of outcomes tested (doubling)	1·33 (0·84-2·18)	0·23

Science at low pre-study odds of true findings

Bottom line

- Small sample sizes
- Modest validation performance at best – with few exceptions
- Unclear role so far for additional prognostic power beyond classic predictors – with few exceptions
- Suspected large reporting bias
- Immense amount of information and undoubtedly great potential for improving performance

The future (?): investigator or data specimen registration

- Upfront study registration has been adopted for randomized clinical trials, as a means for minimizing publication and reporting biases and maximizing transparency
- For molecular research, upfront registration in public of all ideas is counter-intuitive and goes against the individualistic spirit of discovery in basic research
- Instead one could aim for registries of investigators and data specimen collections

Registries of data/sample collections

- Inclusive networks of investigators working on the same disease, set of genes or field
- Promotion of better methods and standardization
- Research freedom for individual participating teams
- Thorough and unbiased testing of proposed hypotheses with promising preliminary data on large-scale comprehensive databases
- Due credit to investigators for both "positive" and "negative" findings
- It is feasible to start from existing coalitions of investigators ("networks") that work on specific diseases, genes or fields

Registries of teams

- The core registry should comprise information on the teams that already participate in a network
- A wider registry should also record all other teams that work on the same field. This should be based on searches of electronic databases (identifying who has published anything on the field of interest), personal contacts, announcement(s) in some major journal(s) and should be an open, evolving process updated at regular intervals
- Depending on the structure and funding opportunities of the existing networks, additional teams may be allowed to join formally and fully in the original network; even if structure or funding considerations do not allow this, additional teams should be simply recorded, so that a picture of the field-at-large is available
- Networks may have qualitative or other pre-requisites for allowing teams to join. These should be developed by the scientists involved, but some central guidance and sharing of experiences would also be useful

Examples of investigator networks: disease-specific

- GENOMOS (osteoporosis)
- GEO-PD (Parkinson's disease)
- Interlymph (lymphoma)
- ILCCO (Lung cancer)
- INHANCE (head and neck cancer)
- Meta-analysis of HIV Host Genetics (HIV)
- WHO craniofacial anomalies consortium (craniofacial anomalies)
- Emerging Risk Factors Collaboration (cardiovascular disease)

Examples of networks: gene- or field-specific

- GSEC (genes involved in environmental carcinogens)
- Web registry of DNA repair genes and cancer
- NIH-US Pharmacogenetics Research Network

How might it look like?

- For cancer X, a network is available with 43 participating teams and with DNA samples on a total of 25000 cases and 27000 controls (total 52000)
- Besides the network, we are also aware of the existence of another 28 teams working on the genetics of this cancer with a total of 18000 cases and 17000 controls (total 35000)
- Promising findings from single teams or findings from meta-analyses of published group data may be tested on a large-scale at the network level
- The certainty for any preliminary finding can be interpreted not only as a function of its statistical significance, but also as a function of the percentage of the total possible evidence upon which it is based; e.g. an odds ratio may have a p-value of 0.001 after 4 teams have tested a specific SNP, but this may be based only on 2600 subjects, i.e. 5% of the total network possible evidence and approximately 3% of the overall possible evidence.
- The network would also ensure that "negative" findings are also disseminated with appropriate credit

What would a network of networks do

- Communication and sharing of expertise in statistical analytical methods, laboratory techniques, practical procedures, logistics of creating and maintaining a network
- Co-ordination of registries, facilitation and avoidance of overlap
- Maximization of efficiency and standardization of methods and procedures
- Electronic list of all registries containing minimal information on all participating teams as well as on non-participating teams
- Eventually keeping updated a "Libro d'oro" of validated molecular information that may be compiled by investigators of each network for the disease/field-at hand

Eventual proposed grading of evidence in molecular research

- III. Single or scattered studies: purely hypothesis-generating, important to register data, regardless of results
- II. Meta-analyses of group data: increasing certainty when several thousand subjects available
- I. Large-scale evidence from individual-level all-inclusive networks: evolving gold standard?

- C. No functional/biological data or negative data
- B. Limited or controversial functional data
- A. Convincing functional data

- 3. No clinical or public health applicability
- 2. Limited applicability
- 1. Clinical/public health applicability

It would not have happened without...

- Thomas A. Trikalinos, MD
- Despina G. Contopoulos-Ioannidis, MD
- Evangelia E. Ntzani, MD
- Fotini B. Karassa, MD
- Georgia Salanti, MSc, PhD
- Anastasia Polycarpou, MD
- Christos Ntais, MD
- Athina Tatsioni, MD
- Panagiotis Kyzas, MD
- Emilios Pakos, MD
- Anna-Bettina Haidich, MSc, PhD
- Elias Zintzaras, MPH, PhD
- Fotini Kavvoura
- Konstantinos Loizou
- Michalis Economou, MD

The issue of indirect comparisons in meta-analysis

Douglas G Altman

Centre for Statistics in Medicine
Oxford, UK

Thanks to Jon Deeks, Anne-Marie Glenny, Fujian Song, Roberto D'Amico, Charlotte Sakarovitch, Mike Bradburn, Alison Eastwood, Julian Higgins

Main question

- **How can we compare the effectiveness of two treatments when there are no randomised trials comparing them?**
- **In many areas randomised controlled trials (RCTs) may not have directly compared specific treatments of interest**
  - e.g. each of two drugs may have been compared to placebo but not with each other directly ('head to head')

Related question

- **Within a meta-analysis, how should we investigate differences between subgroups of studies?**
  - e.g. subgroups defined by patient characteristics, study quality, dose of drug, etc

A hint at the answer

- **In both cases we can make the comparison of interest indirectly via other randomised comparisons**
- **In their simplest form, both questions reduce to this question:**
  - How should we compare two independent estimates of some quantity?
- **This talk will focus on randomised controlled trials (RCTs) - the principles apply to any type of study**

Outline

- **Indirect comparisons**
- **Empirical evidence about direct vs indirect comparisons**
- **Relation to subgroups**
- **Methods of analysis**
- **Assumptions**
- **Conclusions**
  - Whether/when to use the methods

Indirect comparison of B v C Simplest case

Trial	↓	↓
1	A	B
2	A	B
3	A	C
4	A	C
5	A	C

Indirect comparison of B v C
Also 'head-to-head' trials

Trial	↓	↓
1	A	B
2	A	B
3	A	C
4	A	C
5	A	C
6	B	C
7	B	C

7

Indirect comparison of B v C
Also three arm trials

Trial	↓	↓	
1	A	B	
2	A	B	
3	A		C
4	A		C
5	A		C
6		B	C
7		B	C
8	A	B	C
9	A	B	C

8

Indirect comparison of B v C
More than one comparator

Trial	↓	↓	
1	A	B	
2	A	B	
3	A		C
4	A		C
5	A		C
6		B	D
7		B	D
8			C D
9			C D

9

Multiple interventions with same control (e.g. placebo)

Trial	↓	↓		
1	A	B		
2	A	B		
3	A		C	
4	A		C	
5	A		C	
6	A			D
7	A			D
8	A			E
9	A			E

10

Treatments studied in 17 randomised trials reporting 30 day mortality examining thrombolytic therapy for patients with acute coronary syndrome (treatment within six hours of onset). [Hasselblad & Kong 2001]

Trial	Placebo	tPA	APSAC	SK	RPA	Accel tPA
ASSENT	✓	✓				
ECSG	✓	✓				
AIMS	✓		✓			
Bassand1	✓		✓			
GISSI	✓			✓		
ISIS-2	✓			✓		
ISAM	✓			✓		
TEAM-3		✓	✓			
Bassand2		✓	✓			
GISSI-2		✓		✓		
ISG		✓		✓		
ISIS-3		✓	✓	✓		
TIMI-4			✓		✓	
TAPS			✓			✓
INJECT				✓	✓	
GUSTO I				✓		✓
GUSTO III					✓	✓

11

Subset of trials of thrombolytic therapy showing sequential evaluation of new treatments

Trial	Placebo	tPA	APSAC	SK	RPA	Accelerated tPA
ASSENT	✓	✓				
TEAM-3		✓	✓			
ISIS-3			✓	✓		
INJECT				✓	✓	
GUSTO III					✓	✓

12

Indirect comparison

Simplest case

One trial comparing A v B

One trial comparing A v C

We wish to compare B v C

13

Broad approaches

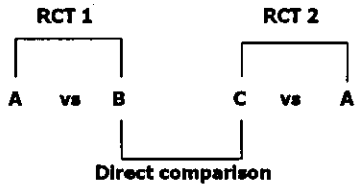
- **Direct comparison of treatments of interest**
  - Ignores randomisation
  - Ignores between study variation
  - Called 'naïve' method
- **Indirect comparison via common comparator**
  - Preserves randomisation
  - Called 'adjusted indirect comparison'

14

Direct/indirect comparison

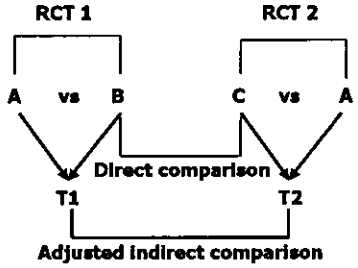
15

Direct/indirect comparison



16

Direct/indirect comparison



17

Direct comparison (naïve method)

- **Treatments for prostatic cancer** [Aro et al 1991]
  - RCT1 compared polyestradiol phosphate (PEP) plus oral ethinyl estradiol (EE) with orchiectomy [BvA]
  - RCT2 compared EE with orchiectomy [CvA]
- **Authors compared PEP+EE with EE as if these two groups had been in the same trial** [BvC]
  - Discarded orchiectomy treatment arm in each trial
- **Observed difference could have been due to various biases, not least the different inclusion criteria in the two trials**

18

Direct comparison (naïve method)

- **Thromboprophylaxis and death after total hip replacement** [Murray et al, 1996]
  - Non pulmonary embolism deaths
  - Pooled data for each treatment from randomised and nonrandomised studies

Prophylaxis	No. of Deaths	No. of Patients	Event rate (%)	95% CI
None	7	3355	0.21	0.08 to 0.43
Heparin	32	10105	0.32	0.22 to 0.44
Warfarin	9	3763	0.24	0.11 to 0.45
Aspirin	1	2649	0.04	0.00 to 0.21
Dextran	6	2618	0.23	0.08 to 0.50

19

Direct comparison (naïve method)

- **"Compared with heparin the use of aspirin caused a significant decrease in non-PE deaths"**
- **"The results must be interpreted with considerable caution, since pooling is based on the assumption that patients and their management are similar in different studies"**
  - Studies were published from 1970 to 1996
  - Seriously understates assumptions
  - Shifts responsibility to the reader!

20

Adjusted indirect comparison: Secondary prevention of stroke [Lowenthal & Buyse, 1994]

	Trials	Deaths
Placebo v Aspirin	7	616
Placebo v Aspirin+dipyridamole	2	284
Aspirin v Aspirin+dipyridamole	2	105

- **8 times as much indirect data as head-to-head data (900 vs 105 deaths)**
- **Main inference based on indirect comparison**
  - Formal statistical comparison of treatment effects in the two sets of trials
  - Cautious interpretation

21

Literature reviews

- **Database of Abstracts of Reviews of Effectiveness (DARE) (1994-1998) was searched for systematic reviews involving meta-analysis of RCTs**
  - both direct and indirect comparisons
  - indirect comparisons alone
- **A systematic search of Medline and other databases to identify published methods for analysing indirect comparisons**

22

Reviews with direct and indirect comparisons

- **734 DARE systematic reviews of RCTs assessed**
- **31/327 (9%) with meta-analyses included some form of indirect comparison**
  - These and 5 others were studied (n=36)
- **13 had direct and indirect comparisons**
  - 10 used adjusted indirect comparison
  - 3 used naïve approach
- **23 had only indirect comparisons**
  - 15 used adjusted indirect comparison
  - 8 used naïve approach
- **Some reviews focused on indirect data and excluded direct comparisons**

23

Reviews with only indirect comparisons

- **Indirect comparison was sometimes done informally**
- **Interpretation was not always appropriate**
  - authors often interpreted results as if direct comparisons had been made

24

Example: Antihypertensive agents for reducing left ventricular hypertrophy [Schmieder et al 1996]

- **39 double blind RCTs with 84 treatment arms (excluded nonRCTs)**

	No. arms
– Diuretics	13
– β-blockers	21
– Calcium channel blockers	19
– ACE inhibitors	18
– Placebo	13

25

Example: Antihypertensive agents for reducing left ventricular hypertrophy [Schmieder et al 1996]

- **39 double blind RCTs with 84 treatment arms (excluded nonRCTs)**

	No. arms	% decrease in left ventricular mass index	
– Diuretics	13	7%	
– β-blockers	21	6%	pairwise
– Calcium channel blockers	19	9%	tests
– ACE inhibitors	18	13%	
– Placebo	13		

- **Conclusions based purely on indirect comparisons**

26

Indirect comparison

- **Focus on the simplest case with three treatments A, B, and C**
  - We have trials comparing A and B and trials comparing A and C
  - We are interested in comparison of B and C
  - Can make the comparison indirectly

27

The adjusted indirect method

- **Trials making different treatment comparison**

  Trial 1 A1 vs B1

  Trial 2 A2 vs C2

- **Comparison between B1 and C2 is adjusted by a common intervention:**

  A1 – B1 vs A2 – C2

- **Extend to multiple trials of each comparison and trials with more than 2 treatment arms**

28

General statistical result

- **Standard error (SE) of the difference between two independent estimates is the square root of the sum of the two SEs**

$$D = X_1 - X_2$$

$$var(D) = var(X_1) + var(X_2)$$
$$SE(D) = \sqrt{[SE(X_1)^2 + SE(X_2)^2]}$$

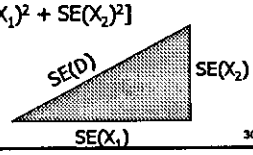
29

General statistical result

- **Standard error (SE) of the difference between two independent estimates is the square root of the sum of the two SEs**

$$D = X_1 - X_2$$

$$var(D) = var(X_1) + var(X_2)$$
$$SE(D) = \sqrt{[SE(X_1)^2 + SE(X_2)^2]}$$



30

The adjusted indirect method
[e.g. Bucher et al, *J Clin Epidemiol* 1997]

T is treatment effect
- e.g. mean difference, log relative risk

T_{AC} is direct comparison between A and B
T_{BC} is direct comparison between A and C
(each based on 1 or more trials)

Adjusted indirect estimate of comparison B v C is

$$T'_{BC} = T_{AC} - T_{AB}$$

and its SE is

$$SE(T'_{AC}) = \sqrt{[SE(T_{AC}) + SE(T_{AB})]}$$

From which we can get confidence interval

31

Indirect comparison as an extension of meta-analysis

- **Meta-analysis using summary data extracted from published studies is a two-stage process**
  - Calculate appropriate summary statistic for each study
  - Calculate the weighted combination of these statistics to provide obtain overall estimate of treatment effect
    - Most standard methods are like this
- **For an indirect comparison we add a third stage**
  - Combine the results of two separate meta-analyses into an overall comparison

32

Other methods for analysing indirect comparisons

- **What other methods can we use to perform an indirect comparison?**

33

Valid methods for indirect comparison

- **Indirect effect of treatment B compared with treatment C can be calculated using:**
  - Difference/ratio of treatment effects **A v B** & **A v C**
    - Derived from separate meta-analyses
  - Logistic regression (for odds ratio only)
  - Meta-regression
    - These methods make slightly different assumptions
  - Complex methods (Bayesian and likelihood-based methods) [e.g. Ades, *Stat Med* 2003]

34

Assumptions

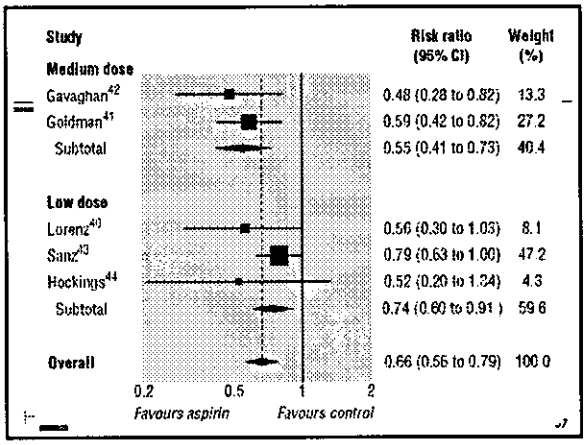
- **All standard meta-analysis assumptions apply**
  - The various studies are all estimating the same effect, such as the effect of A relative to that of B
- **Key additional assumption**
  - Effectively, the choice of comparison is random
  - There are no important differences between the two sets of trials with respect to aspects that could influence (bias) the estimated treatment effect of BvC
- **In practice this may not be correct**
  - There may be reasons for choosing to investigate B or C that are related to patient populations or other study characteristics (or may differ by chance)

35

Example: aspirin after coronary surgery
[Lim et al, *BMJ* 2003]

- **Comparison of low and medium dose aspirin therapy after coronary surgery by using an indirect comparison of placebo-controlled trials, with outcome evaluated by angiography**
- **3 RCTs of low dose aspirin v placebo**
- **2 RCTs of medium dose aspirin v placebo**

36



Example: aspirin after coronary surgery [Lim et al, *BMJ* 2003]

- **Indirect comparison of relative risks (RR)**
  - RR ratio = 0.74 (95%CI 0.52 to 1.06), P = 0.10
- **"Conclusions Medium dose aspirin may more successfully reduce graft occlusion than low dose regimens within the first year after coronary surgery."**
- **Correspondence in *BMJ* considered the possible relation between the time to angiography and the observed effect of aspirin**
  - medium dose angiography at mean of 363 and 367 days
  - low dose at 10, 130, 180

38

Evaluation of different methods

- **Resampling study showed that all the methods evaluated gave very similar answers**
  - ... apart from the naïve method which should never be used
- **Rather than show those results, I will illustrate what happens using a worked example**

39

Event rates for stroke in trials of antithrombotic therapy for patients with atrial fibrillation [Hart et al, 1999]

- **15 RCTs**
- **Each studied 2 or more of 5 treatments**
  - **A Placebo**
  - **B Adjusted dose warfarin**
  - **C Aspirin**
  - **D Low or fixed dose warfarin**
  - **E Low or fixed dose warfarin + aspirin**

40

Event rates for stroke in trials of antithrombotic therapy for patients with atrial fibrillation [Hart et al, 1999]

- **15 RCTs**
- **Each studied 2 or more of 5 treatments**
  - **A Placebo**
  - **B Adjusted dose warfarin**
  - **C Aspirin**
  - **D Low or fixed dose warfarin**
  - **E Low or fixed dose warfarin + aspirin**

41

Event rates for stroke in trials of antithrombotic therapy for patients with atrial fibrillation [Hart et al, 1999]

- **15 RCTs**
- **Each studied 2 or more of 5 treatments**
  - **A Placebo**
  - **B Adjusted dose warfarin**
  - **C Aspirin**
  - **D Low or fixed dose warfarin**
  - **E Low or fixed dose warfarin + aspirin**
- **Main methods investigated**
  - Adjusted indirect comparison (AIC) (simpler cases only)
  - Logistic regression
  - Meta-regression
  - (Mixed models)

42

Trial	Placebo	Adjusted dose warfarin	Aspirin	Low or fixed dose warfarin	Low or fixed dose warfarin + aspirin
	A	B	C	D	E
AFASAK	19/336	9/335	16/336		
SPAF	19/211	8/210	25/552		
BAATAF	13/208	3/212			
CAFA	9/191	6/187			
SPINAF	23/290	7/281			
EAFT i	50/214	20/225	49/230		
ii	40/164		39/174		
ESPS II	23/107		17/104		
LASAF	6/182		5/194		
UK-TIA	8/30		8/34		
SPAF II		19/358	21/357		
AFASAK II		11/170	9/169	14/167	11/171
PATAF		3/131	4/141	4/122	
SPAF III		14/523			48/521
MWNAF	1/153			5/150	

43

Trial	Placebo	Adjusted dose warfarin	Aspirin	Low or fixed dose warfarin	Low or fixed dose warfarin + aspirin
	A	B	C	D	E
AFASAK	19/336	9/335	16/336		
SPAF	19/211	8/210	25/552		
EAFT i	50/214	20/225	49/230		
SPAF II		19/358	21/357		
AFASAK II		11/170	9/169	14/167	11/171
PATAF		3/131	4/141	4/122	
SPAF III		14/523			48/521
MWNAF	1/153			5/150	

44

Trial	Placebo	Adjusted dose warfarin	Aspirin	Low or fixed dose warfarin	Low or fixed dose warfarin + aspirin
	A	B	C	D	E
AFASAK	19/336	9/335	16/336		
SPAF	19/211	8/210	25/552		
EAFT i	50/214	20/225	49/230		
AFASAK II		11/170	9/169	14/167	11/171
PATAF		3/131	4/141	4/122	
SPAF III		14/523			48/521
MWNAF	1/153			5/150	

45

Trial	Placebo	Adjusted dose warfarin	Aspirin	Low or fixed dose warfarin	Low or fixed dose warfarin + aspirin
	A	B	C	D	E
AFASAK	19/336	9/335	16/336		
SPAF	19/211	8/210	25/552		
EAFT i	50/214	20/225	49/230		
AFASAK II		11/170	9/169	14/167	11/171
PATAF		3/131	4/141	4/122	
SPAF III		14/523			48/521
MWNAF	1/153			5/150	

46

Trial	Placebo	Adjusted dose warfarin	Aspirin	Low or fixed dose warfarin	Low or fixed dose warfarin + aspirin
	A	B	C	D	E
AFASAK	19/336	9/335	16/336		
SPAF	19/211	8/210	25/552		
EAFT i	50/214	20/225	49/230		
AFASAK II		11/170	9/169	14/167	11/171
PATAF		3/131	4/141	4/122	
SPAF III		14/523			48/521
MWNAF	1/153			5/150	

47

Atrial fibrillation trials

(a) Indirect comparison

Approach	Method	OR	95% CI	z/t value	P value
AIC	Inverse variance	0.43	(0.22,0.87)	z = –2.36	0.02
AIC	Mantel-Haenszel	0.42	(0.21,0.83)	z = –2.48	0.01
Meta-regression	Weighted linear regression	0.43	(0.23, 0.82)	t = –3.37	0.02
Logistic regression	Fixed effect	0.42	(0.21, 0.83)	z = –2.49	0.01
AIC	DerSimonian and Laird	0.43	(0.21,0.89)	z = –2.28	0.02
Meta-regression	Random effects	0.43	(0.22,0.87)	z = –2.36	0.02
Logistic regression	Random effects	0.41	(0.22, 0.78)	z= 2.71	0.01

48

Atrial fibrillation trials

(b) Combining indirect and direct comparisons

Approach	Method	OR	95% CI	z/t value	P value
AIC	Inverse variance	0.64	(0.40,1.02)	z = −1.86	0.062
AIC	Mantel-Haenszel	0.63	(0.40, 1.01)	z = −1.94	0.05
Meta-regression	Weighted linear regression	0.56	(0.38, 0.83)	t = −2.91	0.004
Logistic regression	Fixed effect	0.62	(0.39, 0.99)	z = −2.00	0.05
AIC	DerSimonian and Laird	0.65	(0.40, 1.04)	z = −1.78	0.07
Meta-regression	Random effects	0.64	(0.40,1.02)	z = −1.86	0.06
Logistic regression	Random effects	0.59	(0.37, 0.93)	z = −2.28	0.02

49

Atrial fibrillation trials

(c) Including multi-arm trials

Approach	Method	OR	95% CI	z/t value	P value
Logistic regression	Fixed effect	0.59	(0.45, 0.79)	z=−3.62	0.0003
Logistic regression	Random effects	0.57	(0.43, 0.75)	z=−3.99	0.0001
Mixed model	Random effects	0.53	(0.37, 0.75)	t=−4.09	0.003

50

Atrial fibrillation trials

(d) All 15 trials

Approach	Method	OR	95% CI	z/t value	P value
Logistic regression	Fixed effect	0.59	(0.45, 0.79)	z=−3.62	0.0003
Logistic regression	Random effect	0.57	(0.43, 0.75)	z=−3.99	0.0001
Mixed model	Random effect	0.53	(0.38, 0.75)	t=−4.09	0.002

51

Subgroups within meta-analyses

- **Subgroups of trials within a meta-analysis may be defined in many ways**
  - Patient factors
  - Study design
  - Methodological quality
- **Subgroups may relate to treatment**
  - Dose
  - Member of a class of drugs
  - Nature of control treatment
- **Same as indirect comparison**

52

Subgroup analyses and indirect comparisons

- **When subgroups define treatments, a formal analysis is exactly the same as an indirect comparison**

53

A common error

- **Comparison of subgroups by means of separate P values**
- **If high dose v placebo: P<0.05**
  and low dose v placebo: P>0.05
  we should not conclude that high dose is more effective
- **Should compare directly the two treatment effects**
  - Indirect comparison

54

Empirical evidence about the validity of indirect comparisons

[Song et al *BMJ* 2003]

- **44 published meta-analyses**
- **Adjusted indirect comparisons usually but not always agreed with the results of head to head randomised trials**
- **The direction of discrepancy between the two estimates was inconsistent**
- **A significant discrepancy (P < 0.05) was seen in three of the 44 comparisons between the direct and the adjusted indirect estimates**

55

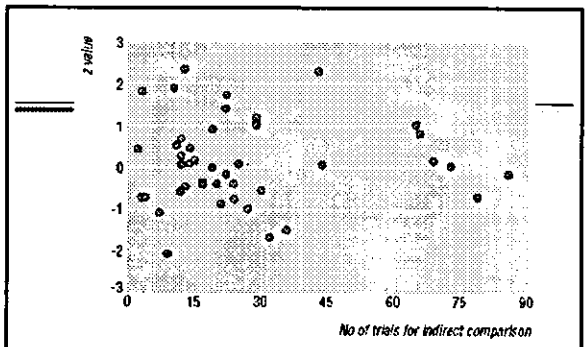


Fig 2 Statistical discrepancy (z value, calculated by dividing difference between direct and indirect estimates by its standard error ($z=\Delta/SE(\Delta)$) and number of trials used in indirect comparison

56

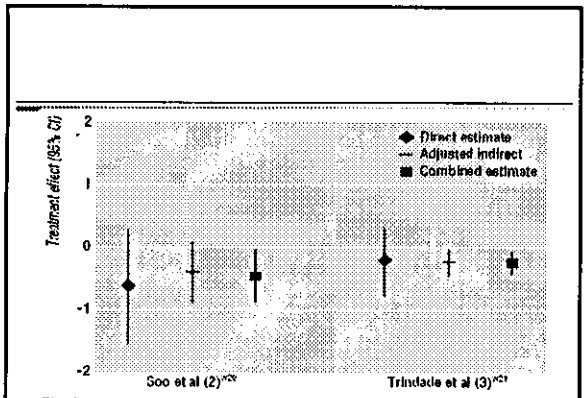


Fig 3 Combination of direct and adjusted indirect estimates in two meta-analyses

57

	Adjusted indirect estimate		
Direct estimate	Significant effect (–) (n=6)	Non-significant effect (n=33)	Significant effect (–) (n=5)
Significant effect (–) (n=8)	5	3	0
Non-significant effect (n=25)	1	23	1
Significant effect (+) (n=11)	0	7	4

*Non-significant effect: difference between intervention groups is non-significant (P>0.05); significant effect (P<0.05) is separated according to whether intervention A is less (–) or more effective (+) than intervention B.

58

Importance of trial similarity: an example (1)

A meta-analysis of paracetamol +codeine versus paracetamol alone in postsurgical pain

	No. of trials	Mean Diff. (95% CI)
Direct	**13**	**6.97 (3.56 to 10.37)**
Adjusted IC	**43**	**-1.16 (-6.95 to 4.64)**

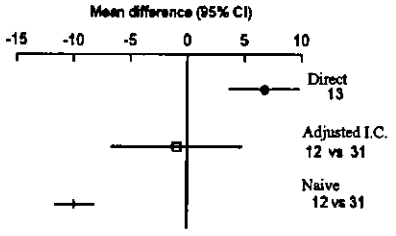
Note: the discrepancy was statistically significant (P=0.02)

Data source: Zhang & Li-Wan-Po (1996)

59

Importance of trial similarity: an example (2)

A meta-analysis of paracetamol +codeine versus paracetamol alone in postsurgical pain



60

Importance of trial similarity: an example (3)

- **Most trials (n=10) in the direct comparison used 600-650 mg paracetamol + 60 mg codeine**
- **Many placebo controlled trials (n=29) used 300 mg paracetamol + 30 mg codeine**
- **When analyse only trials that used 600-650 mg paracetamol + 60 mg codeine the discrepancy was no long statistically significant**

	No. trials	Mean diff (95% CI)
Direct	10	7.28 (3.69 to 10.87)
Adj IC	14	5.72 (-5.37 to 16.81)

61

Conclusions

- **Often no or insufficient direct evidence from randomised trials**
- **Adjusted indirect comparison may provide useful or supplementary information on the relative efficacy of competing interventions**
- **Assumptions not fully clear**
  - Validity depends on the internal validity and similarity of the included trials
  - Exchangeability

62

Can we safely use indirect evidence?

- **Indirect comparisons are non-randomised even though all data are from RCTs**
  - So are subgroup analyses
- **Caution is always needed – the sets of trials may differ in various ways**
  - Cannot establish lack of bias
- **The naïve method should never be used**
- **Appropriate methods of analysis exist and should be used**
- **How reasonable is it to use indirect evidence?**
  - Judge each case on its merits

63

Summary and conclusions

- **When available, results from direct & indirect comparisons usually may not agree but they may not**
- **Essential to consider how similar (sets of) trials are**
  - study and patient characteristics
- **Indirect and direct data should always be considered separately and not automatically pooled**
- **Interpretation should always be cautious in may not view of the observational nature of the data**

64

Some references

Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683-91.

Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy- illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000;21:488-97.

Song F, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326:472-5.

Glenny AM, Altman DG, Song F, et al. Indirect comparison of competing interventions. *Health Technol Assess* (to appear)

65

66

The adjusted indirect method – other terms used in literature

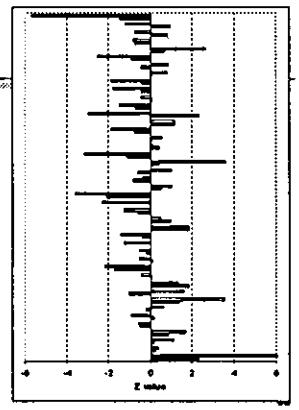
- **Cross-trial comparison**
- **Connected comparative experiment**
- **Network meta-analysis**
- **Mixed comparison**
- **Virtual comparison**

67

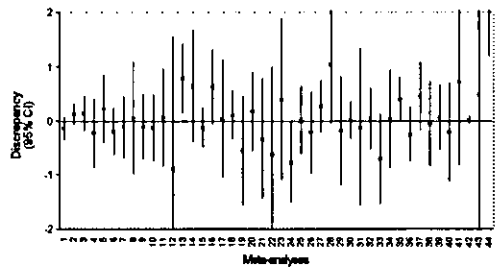
Fig 6.1 (HTA report)

Significant discrepancy ($|Z| \geq$ [illegible]) was observed in 3/43 comparisons between the direct and adjusted indirect estimate (7%); and in 11/43 between the direct and the naïve indirect estimate (26%). The statistical discrepancies between the direct and the naïve indirect estimate are generally greater than that between the direct and the adjusted indirect estimate

Note: . Solid bars - discrepancy between the direct and adjusted indirect estimate. Blank bars - discrepancy between the direct and naïve indirect estimate.



Discrepancy between the direct and the adjusted indirect comparison



69

Methods of comparison and number of significant findings in 44 meta-analyses of competing interventions

	Adjusted I.C		
Direct	Sig(-)	Non-sig	Sig(+)
Sig (-)	5	3	0
Non-sig	1	23	1
Sig(+)	0	7	4

Notes:

1. Non-significant effect: difference between intervention groups is statistically non-significant (P>0.05); significant effect (P≤0.05) is separated according to whether the intervention *A* is less (-) or more effective (+) than intervention *B*.
2. Agreement between the direct and the adjusted indirect estimate: weighted kappa value 0.53.

70

Empirical evidence study: main conclusions

- The naïve indirect comparison should be avoided

- The results of adjusted indirect comparison usually but not always agree with that of head-to-head randomised trials.

- The validity of adjusted indirect comparison depends on the internal validity and similarity of the trials involved.

71

Worked example (1)

This example (data on the next slide) is based on a collaborative overview of RCTs of antiplatelet therapy for the prevention of vascular events in high risk patients (BMJ 1994; 308: 81-106).

Results from 3 RCTs indicated there is no sig difference between the high dose (500-1500 mg/day) and the medium dose (75-325 mg/day) aspirin.

72

Worked example (2)

Comparison	No.of RCTs	Pooled lnRR	SE(lnRR)
High vs Medium	3	-0.0385	0.0902
High vs Control	19	-0.1776	0.0356
Medium vs Control	17	-0.2915	0.0303

73

Worked example (3) - calculation

$$LnRR'_{HvM} = LnRR_{HvC} - lnRR_{MvC}$$

$$= -0.1776 - (-0.2915) = 0.1139$$

and its standard error:

$$SE(lnRR'_{HvM}) = \sqrt{0.0356^2+0.0303^2} = 0.0467$$

Thus, the relative risk of high doses vs medium doses aspirin by the adjusted indirect comparison is: 1.12 (95% CI 1.02 to 1.23), indicating that high doses was less effective than medium doses. This result of indirect comparison should be interpreted with caution. The similarity of patients between high doses trials and medium doses trials needs to be carefully examined.

74

Worked example (4)

Method	lnRR (SE)
Direct comparison:	-0.0385 (0.0902)
Adjusted indirect comparison:	0.1139 (0.0467)

The discrepancy between the direct and the adjusted LC

$$\Delta = lnRR_{Direct} - lnRR_{Indirect} = -0.0385-0.1139=-0.1524$$

$$SE(\Delta) = \sqrt{0.0902^2+0.0467^2} = 0.102$$

Note:
Combining direct and adjusted indirect estimate, the pooled RR = 1.09 (95% CI: 1.00 to 1.18)

75

References

Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683-91.

McAlister F, Laupacis A, Wells G, Sackett D. Users' Guides to the Medical Literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA* 1999;282:1371-7.

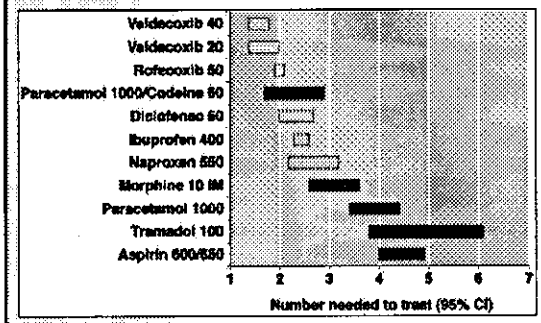
Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy- illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000;21:488-97.

Song F, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326:472-5.

Glenny AM, Altman DG, Song F, et al. Indirect comparison of competing interventions. *Health Technol Assess* (to appear)

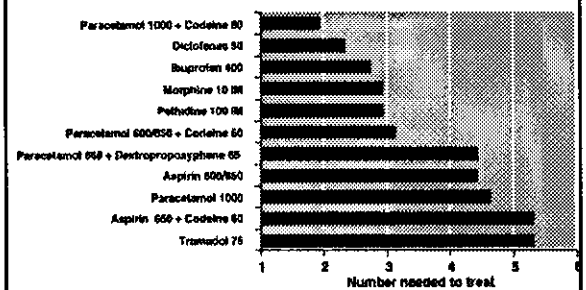
76

Figure 1: League table of number needed to treat (NNT) for at least 50% pain relief over 4-6 hours in patients with moderate to severe pain, all oral analgesics except IM morphine and pethidine and ketorolac



77

League table



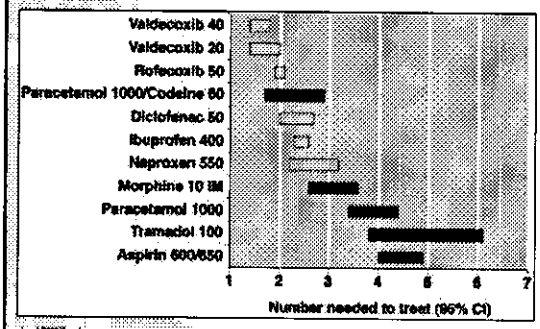
78

Analgesics in acute pain
[Bandolier]

- **"This league table was constructed for analgesics in acute pain. Data were taken from systematic reviews of randomised, double-blind, single-dose studies in patients with moderate to severe pain."**
- **"For each review the outcome was identical - that is at least 50% pain relief over 4-6 hours."**
- **"The league table works because it only has apples, and is not a fruit salad. Only like is compared with like, and there is a common comparator throughout, namely placebo."**

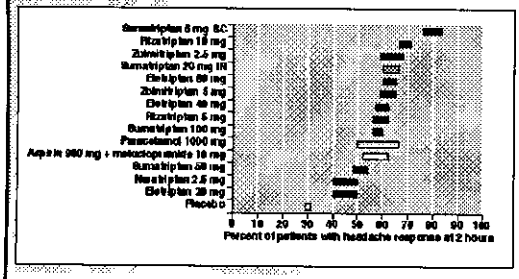
79

Figure 1: League table of number needed to treat (NNT) for at least 50% pain relief over 4-6 hours in patients with moderate to severe pain, all oral analgesics except IM morphine and pethidine and ketorolac



80

Figure 2: Percentage of patients with two hour headache response for each treatment ((bars are 95% confidence interval of the percentage)



81

Trials that can be used to conduct adjusted indirect comparison of Cefuroxime plus Metronidazole versus Co-amoxiclav

Trial	Cefuroxime plus Metronidazole (Cefur-M) SWI/N	Co-amoxiclav (Co-A) SWI/N	Cefotaxime + Metronidazole (Cefot-M) SWI/N
Palmer	2/79	8/69	-
Rowe-Jones	33/454	-	32/453
Kwok	-	7/76	8/88

82

Further trials that can be used to conduct adjusted indirect comparison of Cefuroxime plus Metronidazole vs Co-amoxiclav

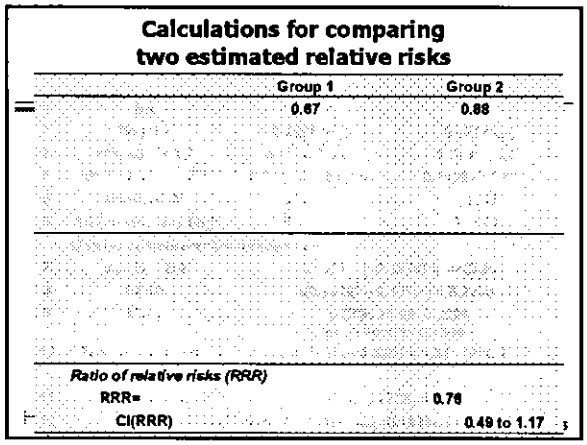
Trials	Cefuroxime + Metronidazole (Cefur-M) SWI/N	Co-amoxiclav (Co-A) SWI/N	Meziocillin SWI/N
Cann [13]	6/52	-	13/43
Diamond [14]	10/53	-	15/51
Stubbs [15]	14/56	-	16/54
Menzies [16]	-	4/30	4/39

83

Calculations for comparing two estimated relative risks

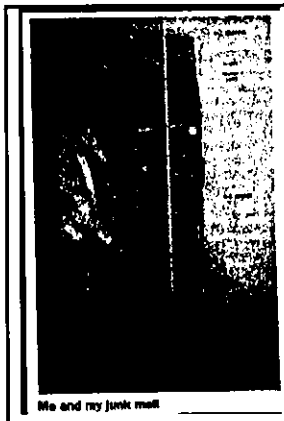
		Group 1	Group 2
1	RR	0.67	0.88
2	*log RR	-0.4005 (E_1)	-0.1278(E_2)
3	95% CI for RR	0.46 to 0.98	0.71 to 1.08
4	*95% CI for log RR	-0.7765 to -0.0202	-0.3425 to 0.0770
5	Width of CI	0.7563	0.4195
6	SE [=width/(2×1.96)]	0.1929	0.1070
	Difference between log relative risks		
7	d (= E_1 - E_2)	-0.4005 - (-0.1278) = -0.2728	
8	SE(d)	$\sqrt{(0.1929^2 + 0.1070^2)}$ = 0.2206	
9	CI(d)	-0.2728 ±1.96×0.2206 or -0.7050 to 0.1598	
10	Test of interaction	z = -0.2726/0.2206 = 1.24 (P=0.2)	
	Ratio of relative risks (RRR)		
11	RRR= exp(d)	exp(-0.2726) = 0.76	
12	CI(RRR)	exp(-0.7050) to exp(0.1598), or 0.49 to 1.17	

84


Calculations for comparing
two estimated relative risks
Group 1
Group 2
0.67
0.88
Ratio of relative risks (RRR)
RRR=
0.76
CI(RRR)
0.49 to 1.17

Graphical displays that might be helpful in interpreting medical data

Ingram Olkin
Stanford University
USA



Montauk L.
Medical Junk Mail.
Lancet 2000;356:344.

Emergency Medicine
San Francisco, USA
Jan 1, 1999 – Dec 31, 1999
122 kg, 2 m high

Outline

(A) Background
-Evidence-based Medicine
-Information Explosion
(B) History and evolution of Meta-Analysis

Information Explosion

1940's:	2,300 biomedical journals
1990's:	**> 25,000**
1940's:	90 psychology journals
1990's:	**> 1,150**
1940's:	90 mathematical journals
1990's:	**> 900**

The *New York Times*
(February 16, 1988) reports:
"The number of scientific articles and journals being published around the world has grown so large that it is starting to confuse researchers, overwhelm the quality control systems of sconce, encourage fraud and distort the dissemination of important findings."

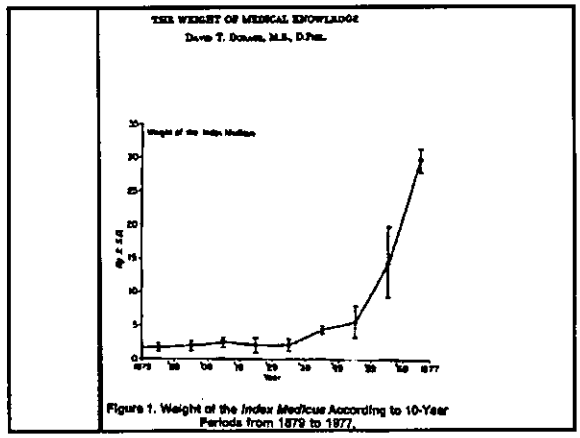
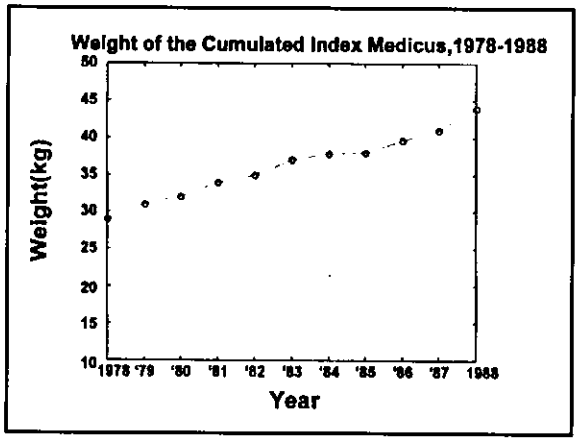


Figure 1. Weight of the *Index Medicus* According to 10-Year Periods from 1879 to 1977.



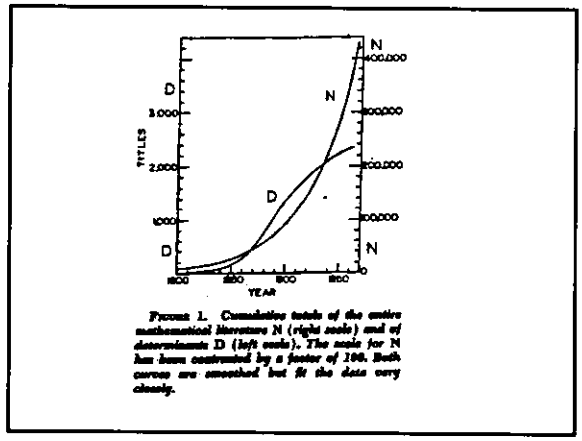
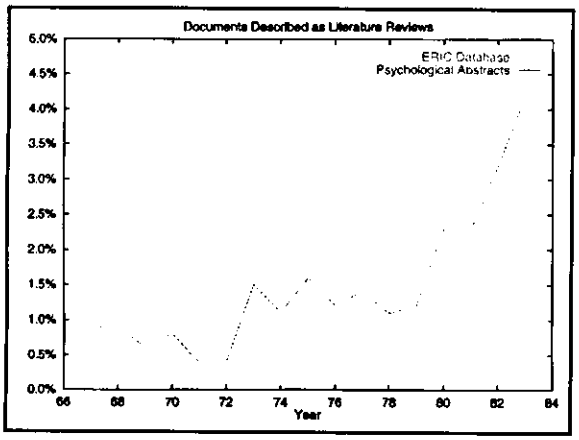


Figure 1. Cumulative totals of the entire mathematical literature N (right scale) and of determinants D (left scale). The scale for N has been contracted by a factor of 100. Both curves are smoothed but fit the data very closely.



MATHEMATICS

900 journals = 3600 issues

10 papers per issue = 35,000 articles

20 papers per issue = 70,000 articles