HTA Tools for Assessing Evidence

Prof. Ken Stein

Peninsula Technology Assessment Group (PenTAG)

UNIVERSITY OF EXETER | MEDICAL SCHOOL

PenTAG

What we're going to cover

- What is evidence?
- Systematic reviews
- Evidence synthesis
- Critical appraisal

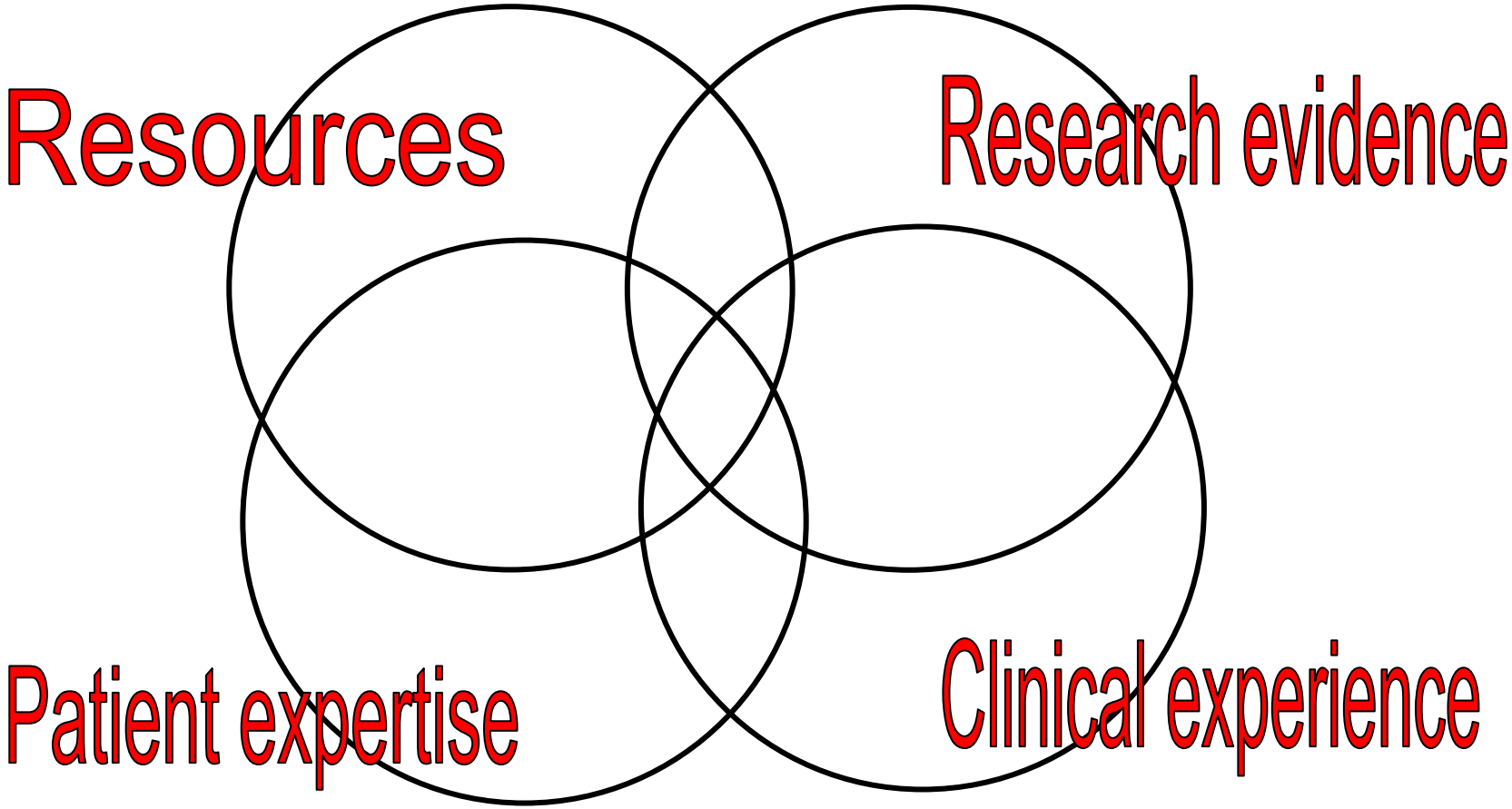
“evidence”

- Help for questions (What will happen in the future?)
  - Does it work?
  - How much will it cost?
  - Will it be good value for money?
  - What impacts will there be on the system?

- Best evidence?
  - Depends on
    - Question
    - Potential for biased answer

How does evidence fit into decision making...?

What factors influence a *policy* decision to adopt a health technology?


Resources
Research evidence
Patient expertise
Clinical experience

Scope for a technology assessment

- Population
- Intervention
- Comparator
- Outcome

- PICO

- Note problems with scoping in a rigid system

- Does the health technology produce a useful health gain (does it work?)
  - Efficacy …
  - Effectiveness …
- Which outcomes are important?
- For whom?
- Instead of what? What will be replaced (if anything) by the new technology?

- So, how do we tell whether a technology produces an effect?

- RCTs
- Systematic review

“Hierarchy of Evidence”

...what should we believe?

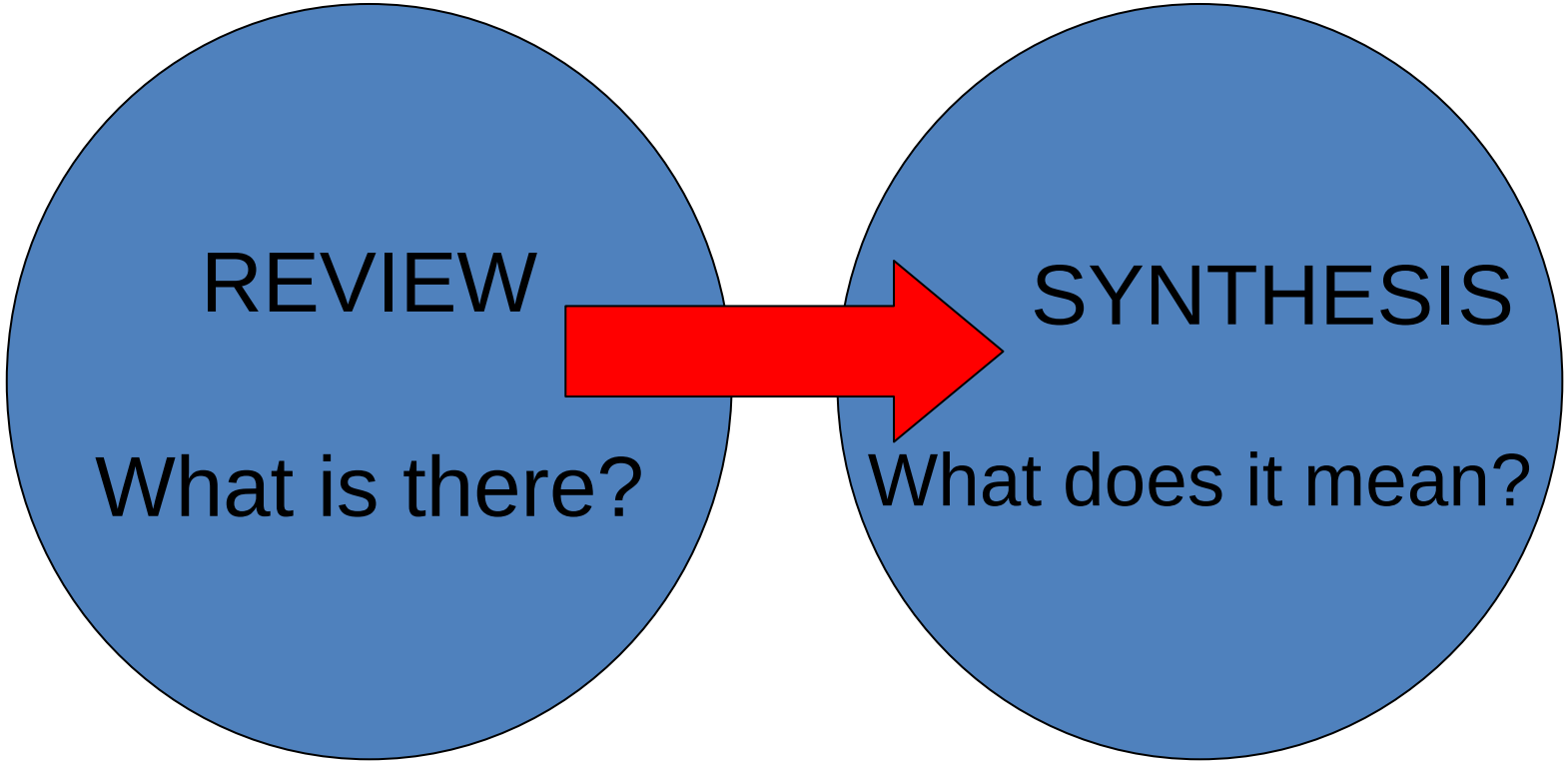
Level	Description	Example
One	Strong evidence from at least one systematic review of well designed randomised controlled trials (RCTS)	Meta-analyses The Cochrane Collaboration
Two	Evidence from at least one properly designed RCT of appropriate size	Articles published in peer-reviewed journals
Three	Evidence from well designed trials without randomization: cohort, time series or matched case controlled studies	Articles published in peer-reviewed journals
Four	Evidence from well designed non-experimental studies from more than one centre or research group	Articles published in peer-reviewed journals
Five	Opinions from respected authorities, based on clinical evidence, descriptive studies or reports from committees	NICE guidelines Evidence-based local procedures and care pathways
Six	Views of colleagues/peers	Clinical colleagues or members of the multidisciplinary team

Systematic review and Evidence Synthesis

Review - the process of bringing together a body of evidence from different sources

Synthesis - stage of a review in which evidence extracted from different sources is *juxtaposed* to identify patterns & direction in the findings, or *integrated* to produce an overarching, new explanation/theory which attempts to account for the range of findings

Prof. Nick Mays, LSHTM, 2006


REVIEW
What is there?
SYNTHESIS
What does it mean?

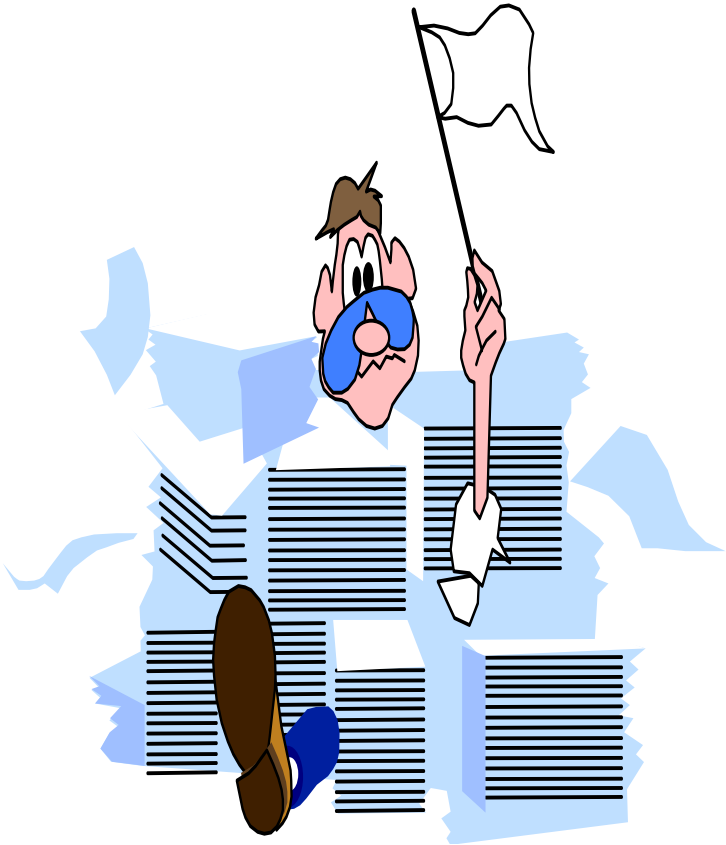
First Systematic Review?

"As it is no easy matter to root out prejudices ... it became a requisite to exhibit a full and impartial view of what had hitherto been published on the scurvy ... by which the sources of these mistakes may be detected. Indeed, before the subject could be set in a clear and proper light, it was necessary to remove a great deal of rubbish."

James Lind, 1753

why bother?
practical perspective

- *Two Million* articles per year in 20,000 biomedical journals
- BMJ and NEJM
  - 4,400 pages
  - 1,100 articles
  - IN ONE YEAR (1992)



“SYSTEMATIC”?

“Reviews are defined as systematic when the account of the search, appraisal and synthesis methods (to minimise bias and random errors) would, in theory, permit the replication of the review by others”

NHS HTA Programme

“SYSTEMATIC”

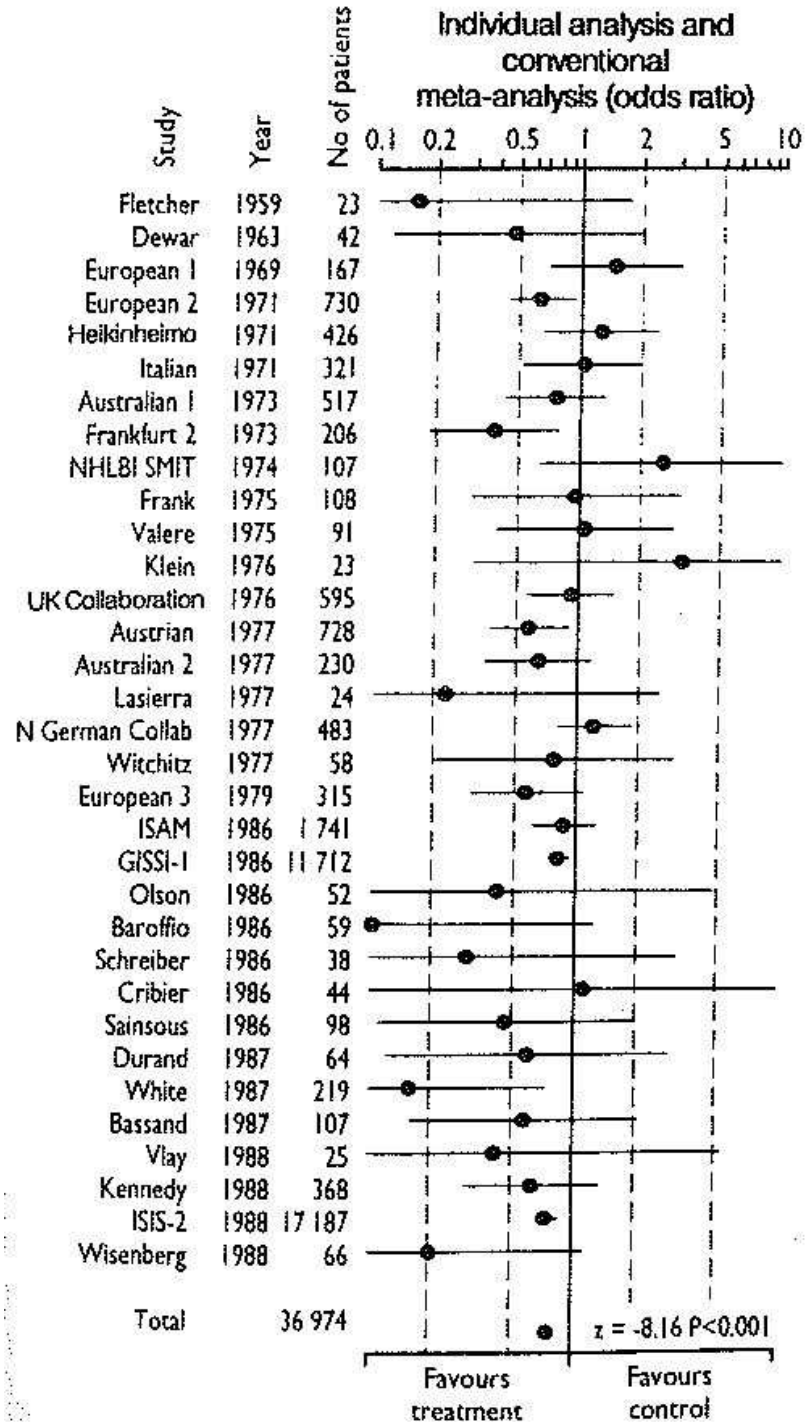
=

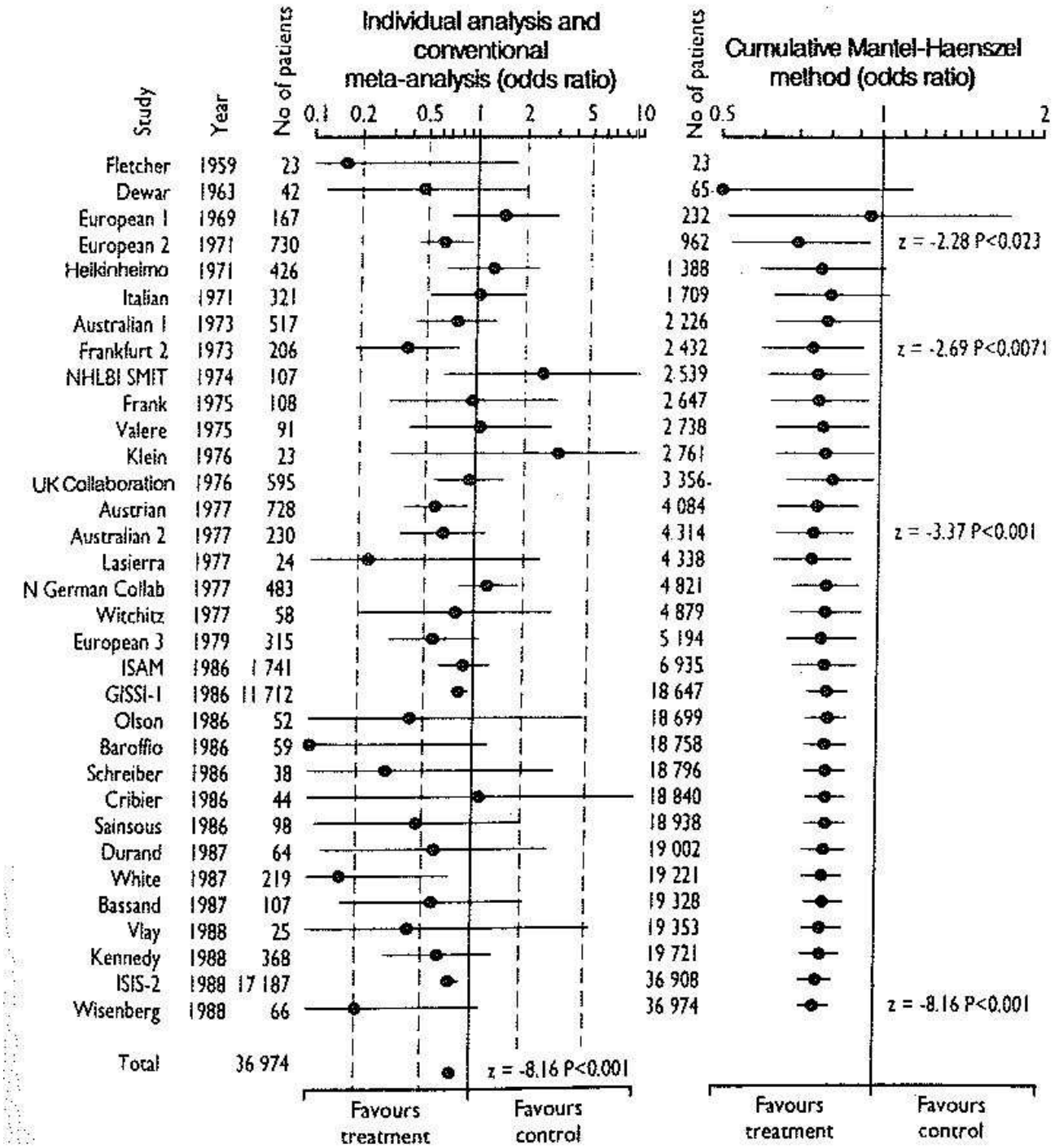
TRANSPARENCY ABOUT METHODS TO MINIMISE BIAS

“The clinical value of thrombolysis

... remains uncertain.”

Oxford Textbook of Medicine, 1987


Individual analysis and conventional meta-analysis (odds ratio)
Study
Year
No of patients
0.1 0.2 0.5 1 2 5 10
Fletcher 1959 23
Dewar 1963 42
European 1 1969 167
European 2 1971 730
Heikinheimo 1971 426
Italian 1971 321
Australian 1 1973 517
Frankfurt 2 1973 206
NHLBI SMIT 1974 107
Frank 1975 108
Valere 1975 91
Klein 1976 23
UK Collaboration 1976 595
Austrian 1977 728
Australian 2 1977 230
Lasierra 1977 24
N German Collab 1977 483
Witchitz 1977 58
European 3 1979 315
ISAM 1986 1 741
GISSI-1 1986 11 712
Olson 1986 52
Baroffio 1986 59
Schreiber 1986 38
Cribier 1986 44
Sainsous 1986 98
Durand 1987 64
White 1987 219
Bassand 1987 107
Vlay 1988 25
Kennedy 1988 368
ISIS-2 1988 17 187
Wisenberg 1988 66
Total 36 974
z = -8.16 P<0.001
Favours treatment
Favours control

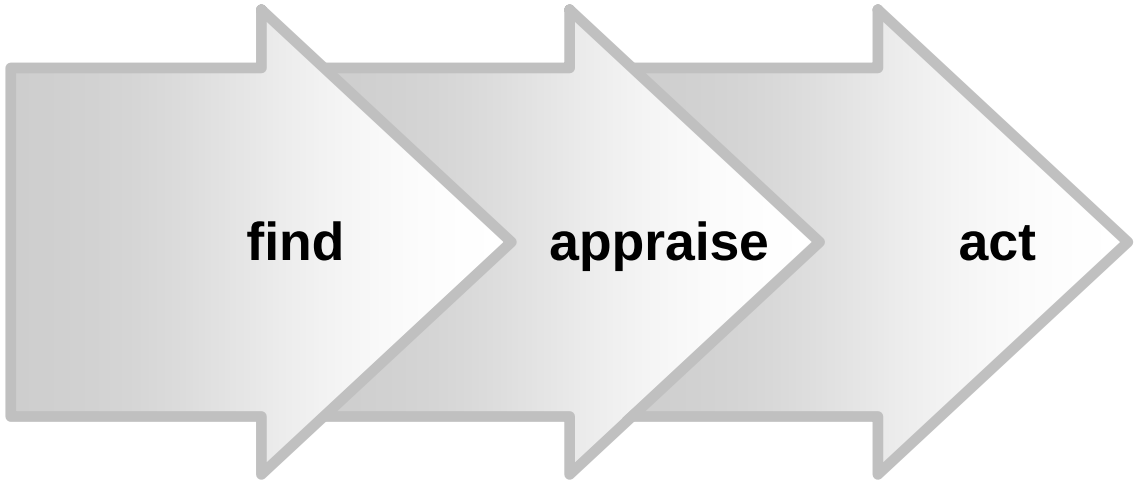

Study
Year
No of patients
Individual analysis and conventional meta-analysis (odds ratio)
0.1 0.2 0.5 1 2 5 10
Cumulative Mantel-Haenszel method (odds ratio)
0.5 1 2
Fletcher 1959 23 23
Dewar 1963 42 65
European 1 1969 167 232
European 2 1971 730 962
z = -2.28 P<0.023
Heikinheimo 1971 426 1 388
Italian 1971 321 1 709
Australian 1 1973 517 2 226
Frankfurt 2 1973 206 2 432
z = -2.69 P<0.0071
NHLBI SMIT 1974 107 2 539
Frank 1975 108 2 647
Valere 1975 91 2 738
Klein 1976 23 2 761
UK Collaboration 1976 595 3 356
Austrian 1977 728 4 084
Australian 2 1977 230 4 314
z = -3.37 P<0.001
Lasierra 1977 24 4 338
N German Collab 1977 483 4 821
Witchitz 1977 58 4 879
European 3 1979 315 5 194
ISAM 1986 1 741 6 935
GISSI-1 1986 11 712 18 647
Olson 1986 52 18 699
Baroffio 1986 59 18 758
Schreiber 1986 38 18 796
Cribier 1986 44 18 840
Sainsous 1986 98 18 938
Durand 1987 64 19 002
White 1987 219 19 221
Bassand 1987 107 19 328
Vlay 1988 25 19 353
Kennedy 1988 368 19 721
ISIS-2 1988 17 187 36 908
Wisenberg 1988 66 36 974
z = -8.16 P<0.001
Total 36 974
z = -8.16 P<0.001
Favours treatment
Favours control
Favours treatment
Favours control

Inputs to HTA reviews

- Search strategies – long, detailed, complex
- Importance of information scientists
  - Framing searches
  - Using different databases
- Searches for NICE appraisals are more extensive than for Cochrane reviews

Steps in a systematic review
(or – have they done a useful one?)

- Ask the question
- Get the evidence
  - SEARCH
  - HOW?
- Check the stuff
  - APPRAISAL
  - HOW?
- Synthesise the stuff
  - what does it mean?
  - meta-analysis (quantitative data)
- Apply in the decision



Critical appraisal tools for evidence

- To add…
- JAMA & CASP
- What is used in PenTAG TAR and other TARs
- Summary – what they all do = risk of bias (Cochrane)

Evidence synthesis

- Bringing together all the forms of evidence
- Meta-analysis (MA, MR & MTC)
- Narrative synthesis
- Qualitative synthesis

- *Do we want to look at examples of all of them? Could have (ideally) examples of each to hand for discussion…*

HTA approaches to evidence synthesis

- STRUCTURE
- Number of reviewers (outline qualifications and training with list of UK courses)
- Information science (outline qualifications and trawling vs spear fishing)
- Note different systems e.g. HTACG vs NICE
- NICE has Decision Support Unit (useful for high throughput complex topics – note some recent titles)

HTA approaches to evidence synthesis

- PROCESS
- Critical appraisal tools (simple)
- Meta-analysis
- Meta-regression
- Multiple treatment comparisons

HTA approaches to evidence synthesis

- OUTCOME
- How detailed does an assessment of evidence quality and summary effect need to be?
  - Think about it for your system (SMALL GROUPS – 5 OR 10 MINS?)
- How much does it cost?
- Note academic gains (public good with open publication)

Multiple treatment comparison

- Many studies compare a treatment only to placebo
- We can use these to infer what the *comparative effectiveness* may be against an active comparator
- Multiple Treatment Comparison or indirect comparison

Network of Evidence

- We have trials of:
  - Drug A vs placebo (odds ratio, OR)
  - Drug B vs placebo (OR)

- Then OR of Drug A vs Drug B:

$$\text{Ln OR}(A\text{-}B) = \ln \text{OR}\ (A\text{-}P) + \ln \text{OR}\ (B\text{-}P)$$

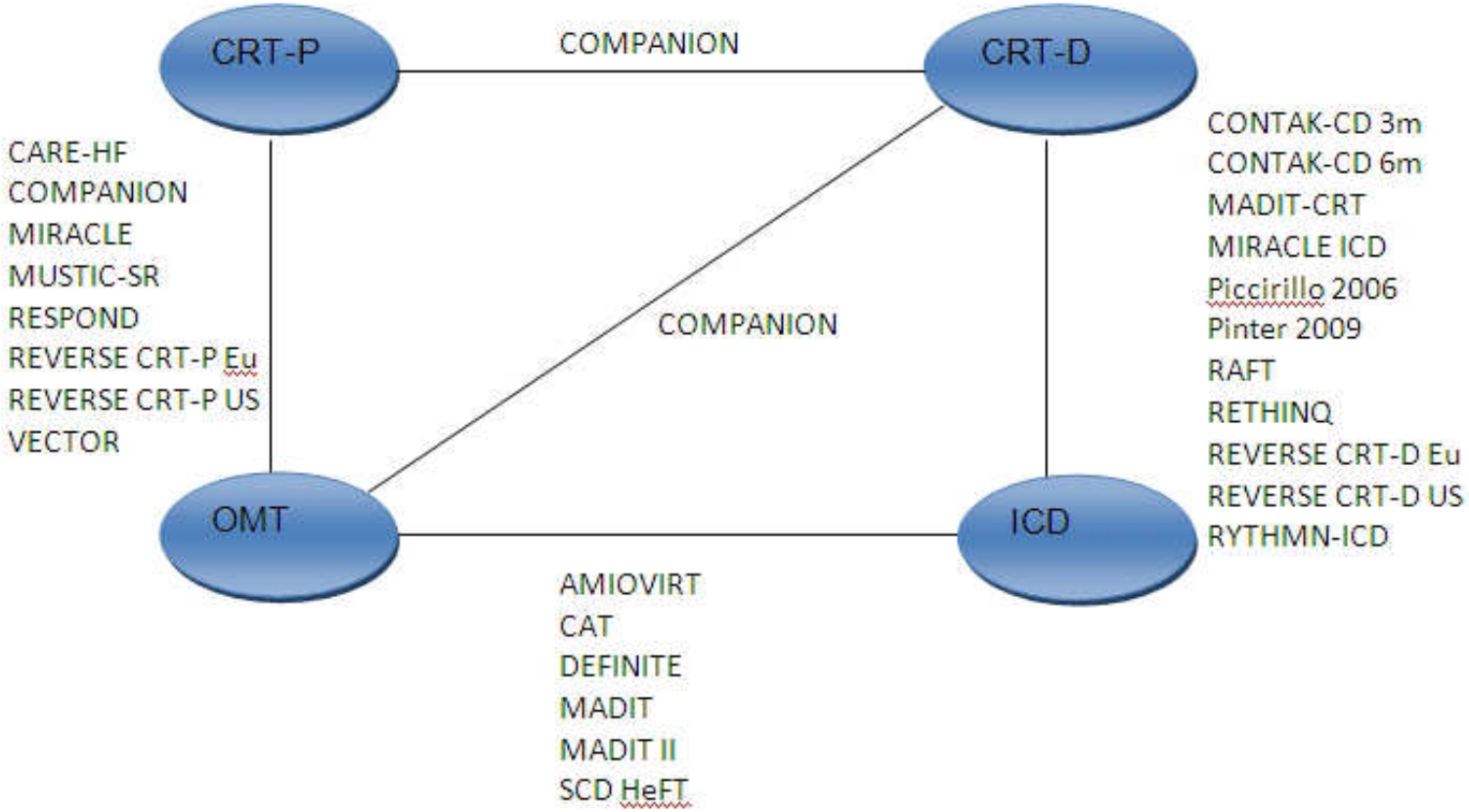
Depending on *exchangeability* i.e. are the populations and placebo interventions sufficiently similar?

Can explore exchangeability statistically. Similar to heterogeneity

Assumptions in MTC are pretty much the same as those used in meta-analysis

MTC allows direct AND indirect evidence to be combined

Network of RCTs



Critical appraisal

- Your health ministry has received a request to fund minimally invasive surgery for benign prostatic enlargement

- What is the question?

Problem and resource

- In men experiencing the symptoms of benign prostatic enlargement, do minimally invasive techniques produce better outcomes than TURP in terms of symptoms and complications?

- You have identified this review
  - *Minimally invasive treatments for benign prostatic enlargement: systematic review of randomised controlled trials. BMJ 2008;337:a1662*

Steps in appraising areview

- What’s the question?
- Did they identify the relevant articles?
- Did they select for inclusion well enough?
- Were the included studies valid?
- Were the results similar?
- What were the results?

What question (PICO) did the systematic review address?

What is best?	Where do I find the information?
The main question being addressed should be clearly stated. The exposure, such as a therapy or diagnostic test, and the outcome(s) of interest will often be expressed in terms of a simple relationship.	The Title, Abstract or final paragraph of the Introduction should clearly state the question. If you still cannot ascertain what the focused question is after reading these sections, search for another paper!

Is it unlikely that important, relevant studies were missed?

What is best?	Where do I find the information?
The starting point for comprehensive search for all relevant studies is the major bibliographic databases (e.g., Medline, Cochrane, EMBASE, etc) but should also include a search of reference lists from relevant studies, and contact with experts, particularly to inquire about unpublished studies. The search should not be limited to English language only. The search strategy should include both MESH terms and text words.	The Methods section should describe the search strategy, including the terms used, in some detail. The Results section will outline the number of titles and abstracts reviewed, the number of full-text studies retrieved, and the number of studies excluded together with the reasons for exclusion. This information may be presented in a figure or flow chart.

Were the criteria used to select articles for inclusion appropriate?

What is best?	Where do I find the information?
The inclusion or exclusion of studies in a systematic review should be clearly defined a priori. The eligibility criteria used should specify the patients, interventions or exposures and outcomes of interest. In many cases the type of study design will also be a key component of the eligibility criteria.	The Methods section should describe in detail the inclusion and exclusion criteria. Normally, this will include the study design.

Were the included studies sufficiently valid for the type of question asked?

What is best?	Where do I find the information?
The article should describe how the quality of each study was assessed using predetermined quality criteria appropriate to the type of clinical question (e.g., randomization, blinding and completeness of follow-up)	The Methods section should describe the assessment of quality and the criteria used. The Results section should provide information on the quality of the individual studies.

Were the results similar from study to study?

What is best?	Where do I find the information?
Ideally, the results of the different studies should be similar or homogeneous. If heterogeneity exists the authors may estimate whether the differences are significant (chi-square test). Possible reasons for the heterogeneity should be explored.	The Results section should state whether the results are heterogeneous and discuss possible reasons. The forest plot should show the results of the chi-square test for heterogeneity and if discuss reasons for heterogeneity, if present.

What are the results?

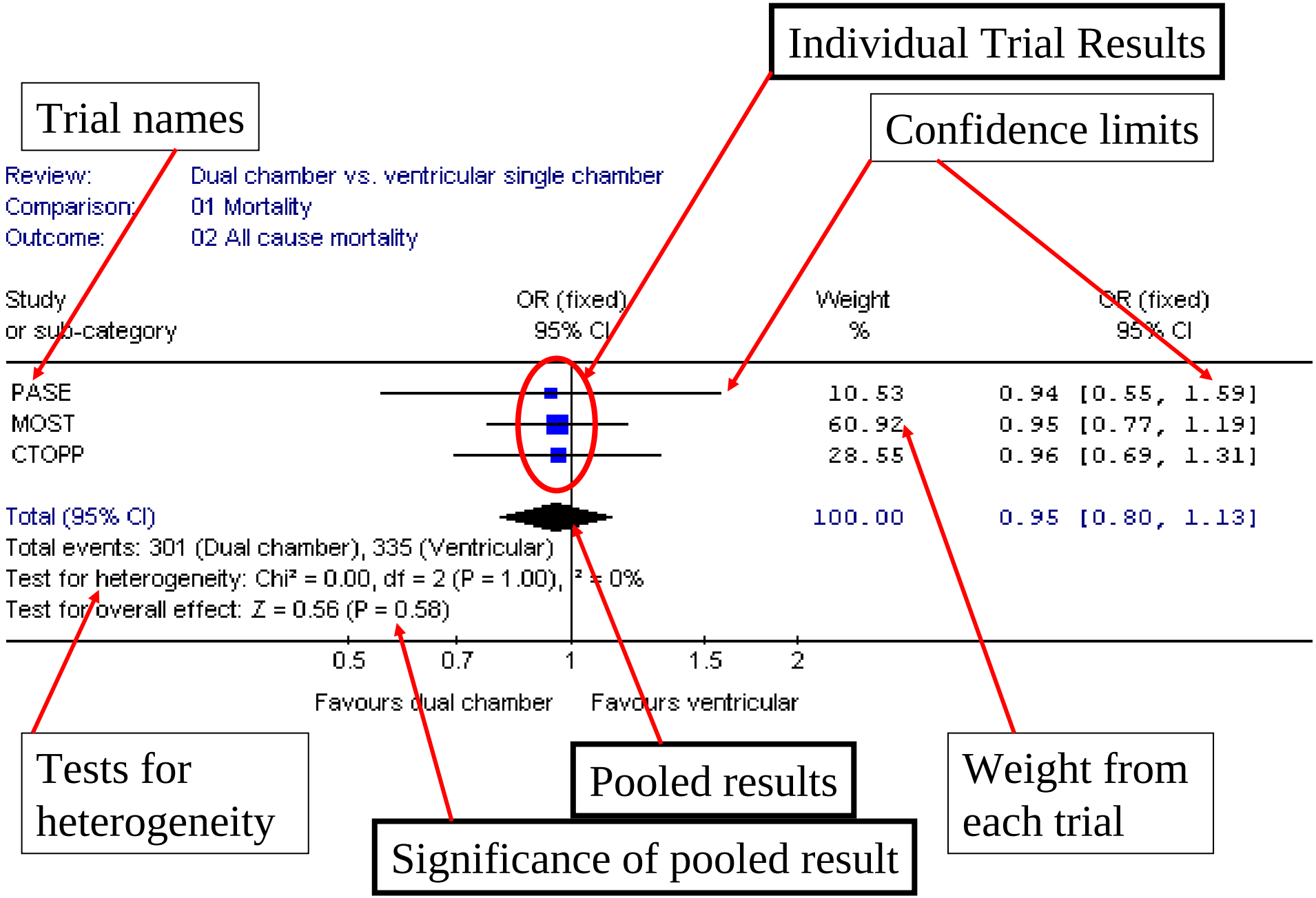
Did they carry out a meta-analysis? Was this appropriate? This depends on the degree of clinical and statistical heterogeneity.

CLINICAL – consider whether the interventions, comparators and outcomes of the included studies are comparable. If not, then it may not be wise to combine the results of studies. You may see that outcome measures have been converted, or standardised, so that they are presented in terms of a proportion of standard deviation for that method. Although this provides a way of considering whether there is an effect, it makes interpretation of the size of any effect that is shown difficult.

STATISTICAL – there will be a test for heterogeneity and, perhaps, the I^2 statistic. The former is a test of the hypothesis that the studies are giving results that are different beyond what you might expect by chance if they are all looking at the same underlying effect. This is a low power test and usually there are no that many studies, so a level of 0.10 is usually taken as demonstrating significant heterogeneity. The I^2 statistic gives the proportion of the variation in the results of the trials that is due to heteregeneity. Levels of 50% are usually taken as high.

What to do about heterogeneity? Well, you can (a) split the studies up so that there are more homogenous groups; (b) exclude studies that seem to be accounting for the heterogeneity; (c) use a random effects model for the analysis. This relaxes the assumption that all the studies are examining a single underlying quantity, but that the underlying quantity itself may vary. Using a random effects model does not make the heterogeneity go away, but it means that more uncertainty is taken into account. From a practical point of view this means that the confidence intervals from a random effects model are wider than from a fixed effect model

If no meta-analysis has been done, ask whether the authors have examined the variation in the studies and drawn careful conclusions about what the direction and size of effects may be within the group of studies that have been reviewed.


Individual Trial Results
Trial names
Confidence limits
Review: Dual chamber vs. ventricular single chamber
Comparison: 01 Mortality
Outcome: 02 All cause mortality
Study or sub-category
OR (fixed) 95% CI
Weight %
OR (fixed) 95% CI
PASE 10.53 0.94 [0.55, 1.59]
MOST 60.92 0.95 [0.77, 1.19]
CTOPP 28.55 0.96 [0.69, 1.31]
Total (95% CI) 100.00 0.95 [0.80, 1.13]
Total events: 301 (Dual chamber), 335 (Ventricular)
Test for heterogeneity: Chi² = 0.00, df = 2 (P = 1.00), I² = 0%
Test for overall effect: Z = 0.56 (P = 0.58)
0.5
0.7
1
1.5
2
Favours dual chamber
Favours ventricular
Tests for heterogeneity
Pooled results
Weight from each trial
Significance of pooled result

Decision…

- Should we conclude that minimally invasive techniques to treat benign prostatic enlargement are clinically effective?
- Is there anything else you would like to know before agreeing to fund such treatments?