جامعة الملك سعود
King Saud University

Evidence-based Health Care (2)

Dr. Nada A. AlYousefi

Associate Professor, Postgraduate Trainer and Consultant of Family Medicine
International Board Certified Lactation Consultant
Department of Family and Community Medicine
College of Medicine, King Saud University (KSU)

nalyousefi@ksu.edu.sa

January 2019

Introduction to Critical Appraisal

- **WHY APPRAISE THE EVIDENCE?**
  - Where an article is published, or who wrote it should not be an indication of its trustworthiness and relevance. Using critical appraisal skills and tools enables users of research evidence to reach their own judgements.


Paper
Paper
Paper

EDUCATION

Harvard Finds Scientist Guilty of Misconduct

By NICHOLAS WADE AUG. 20, 2010

Harvard University said Friday that it had found a prominent researcher, Marc Hauser, "solely responsible" for eight instances of scientific misconduct.

Hours later, Dr. Hauser, a rising star for his explorations into cognition and morality, made his first public statement since news of the inquiry emerged last week, telling The New York Times, "I acknowledge that I made some significant mistakes" and saying he was "deeply sorry for the problems this case had caused to my students, my colleagues and my university."

Dr. Hauser is a leader in the field of animal and human cognition, and in 2006 wrote a well-received book, "Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong." Harvard's findings against him, if sustained, may cast a shadow over the broad field of scientific research that depended on the particular research technique often used in his experiments.

Harvard itself had faced growing criticism for not releasing more details of

They concluded that Hauser had fabricated data in one study, manipulated results in multiple experiments, and incorrectly described how studies were conducted.

Importance

- Combat **information overload.**
- Identify papers that are **clinically relevant.**
- **Continuing Professional Development (CPD)** - critical appraisal is a requirement for the evidence based medicine component of many membership exams.
- Research interest.

How to read a research article?

Structure of an article

1. Title
2. Abstract
3. Introduction
4. Background / review of literature
5. Organizational context
6. Methodology
7. Results
8. Discussion

Structure of an article

1.Title

Not always a good indication of the content of the article

Example: *"The Risks of Autonomy: Empirical Evidence for the Necessity of a Balance Management in Promoting Organizational Innovativeness" ??????*

Structure of an article

2. Abstract

Sometimes unclear. What should be in it:

a summary of the the research question, key methods, results and conclusions of the study

Structure of an article

3. Introduction

Should contain the research question (PICO) or hypotheses tested

4. Background / review of literature

Research questions occur in the context of an already-formed body of knowledge. The background should address this context, help set the rationale for the study, and explain why the questions being asked are relevant.

Structure of an article

5. Research setting (organizational context)

6. Methodology

Should describe exactly how the research was carried out:

- **Sample:** characteristics, selection, number, non-response
- **Measures:** description of tests / questionnaires (validated?), data, outcome measures
- **Procedure:** study design (qualitative, quantitative, controlled?)

7. Results

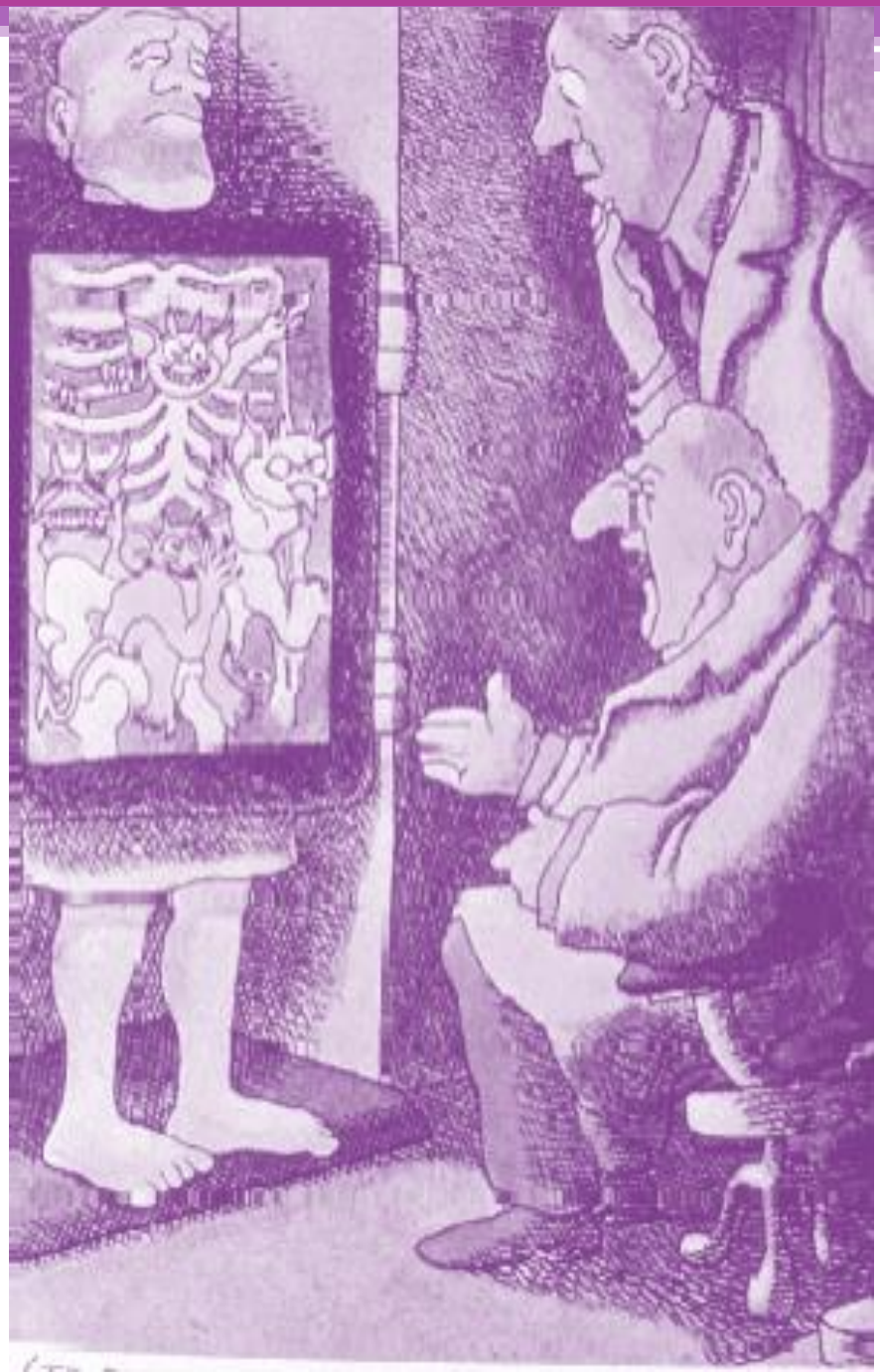
Should tell the reader what the findings were. All outcome measures must be reported and confidence intervals for effect sizes should be presented.

8. Discussion

- Interpretation of the results / relation to theory
- Comparison with the results of other studies
- Weaknesses / limitations of the study
- Implications
- Recommendations

In general

- Don’t let yourself be taken in by scientific jargon and complex use of language. Good articles are written in plain English!
- Even authorative journals with a high impact factor contain bad articles and vice versa.
- Focus on research question, study design and outcome.
- Be critical!! Always ask yourself: does this make sense?



Critical appraisal of different study designs

- To critically appraise a journal article, you would have to start by assessing the research methods used in the study.
- **Checklists** - specific to the study design.
- The following checklists are commonly used:
  - **CEBMH** http://cebmh.warne.ox.ac.uk/cebmh/education_critical_appraisal.htm

Tools for critical appraisal

- CASP: Critical Appraisal Skills Programme Checklists
- Critically Appraised Topics: generic systematic reviews (ACP Journal club)
- SIGN: Scottish Intercollegiate Guidelines Network
- GATE Frame

Critical Appraisal Skills Programme (CASP)

Making sense of evidence

CASP CHECKLISTS

This set of eight critical appraisal tools are designed to be used when reading research, these include tools for Systematic Reviews, Randomised Controlled Trials, Cohort Studies, Case Control Studies, Economic Evaluations, Diagnostic Studies, Qualitative studies and Clinical Prediction Rule.

CASP Checklists (click to download)

CASP Systematic Review Checklist	CASP Qualitative Checklist
CASP Randomised Controlled Trial Checklist	CASP Case Control Checklist
CASP Diagnostic Checklist	CASP Cohort Study Checklist
CASP Economic Evaluation Checklist	CASP Clinical Prediction Rule Checklist

Critical Appraisal

Print this Page

These checklists will help you work through the process of critically appraising a research paper.

Overviews	- to appraise systematic reviews and meta analysis	download
Treatment	- to appraise single randomised controlled trials	download
Diagnosis	- to appraise studies of diagnosis	download
Prognosis	- to appraise studies of prognosis	download

CEBM

CENTRE FOR EVIDENCE-BASED MEDICINE

HOME EDUCATION & TRAINING EVENTS EBM RESOURCES

Critically Appraising the Evider

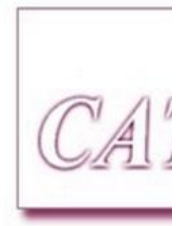
Evaluation a report of a study to determine whether it is valid, important and applicable to your c

Home > EBM Resources > Tools > Critically Appraising the Evidence

Critical Appraisal tools

Critical appraisal worksheets to help you appraise the reliability,

CATMaker ar



Put the "like" into your

Critical Appraisal Worksheets

English

- Systematic Review Critical Appraisal Sheet
- Diagnosis Critical Appraisal Sheet
- Prognosis Critical Appraisal Sheet
- Therapy / RCT Critical Appraisal Sheet

German – Translated by Johannes Pohl and Martin Sadilek

- Systematic Review German Translation (PDF)
- Diagnosis German Translation (PDF)
- Prognosis German Translation (PDF)
- Therapy / RCT German Translation (PDF)

Spanish – Translated by Ana Cristina Castro

- Systematic Review (PDF)
- Diagnosis (PDF)
- Prognosis Spanish Translation (PDF)
- Therapy / RCT Spanish Translation (PDF)

Lithuanian – Translated by Tumas Beinortas

- Systematic review appraisal Lithuanian (PDF)
- Diagnostic accuracy appraisal Lithuanian (PDF)
- Prognostic study appraisal Lithuanian (PDF)
- RCT appraisal sheets Lithuanian (PDF)

5-step approach

EBM is a 5-step approach

1. Formulate an answerable question (PICOC)
2. Search for the best available evidence
3. Critically appraise the quality of the found evidence
4. Integrate the evidence with managerial expertise and organizational concerns and apply
5. Monitor and evaluate the results

APPRAISING THE EVIDENCE

- **1. Is the study valid?**
  - **Decide whether studies have been undertaken in a way that makes their findings reliable.**
- **2. What are the results?**
  - **Make sense of the results.**
- **3. Are the results useful?**
  - **Know what these results mean in the context of the decision that needs to be made.**

What's A Paper on Therapy? Randomised Control Trials

- **Objectives**
  - **Understand why randomised controlled trials produce the most reliable evidence for questions about effectiveness**
  - **Understand the important elements of trial design to minimise bias**
  - **Have critically appraised a randomised controlled trial**

What's A Paper on Therapy?

- Clinical Trial (Controlled) Compares

INTERVENTION
with
CONTROL

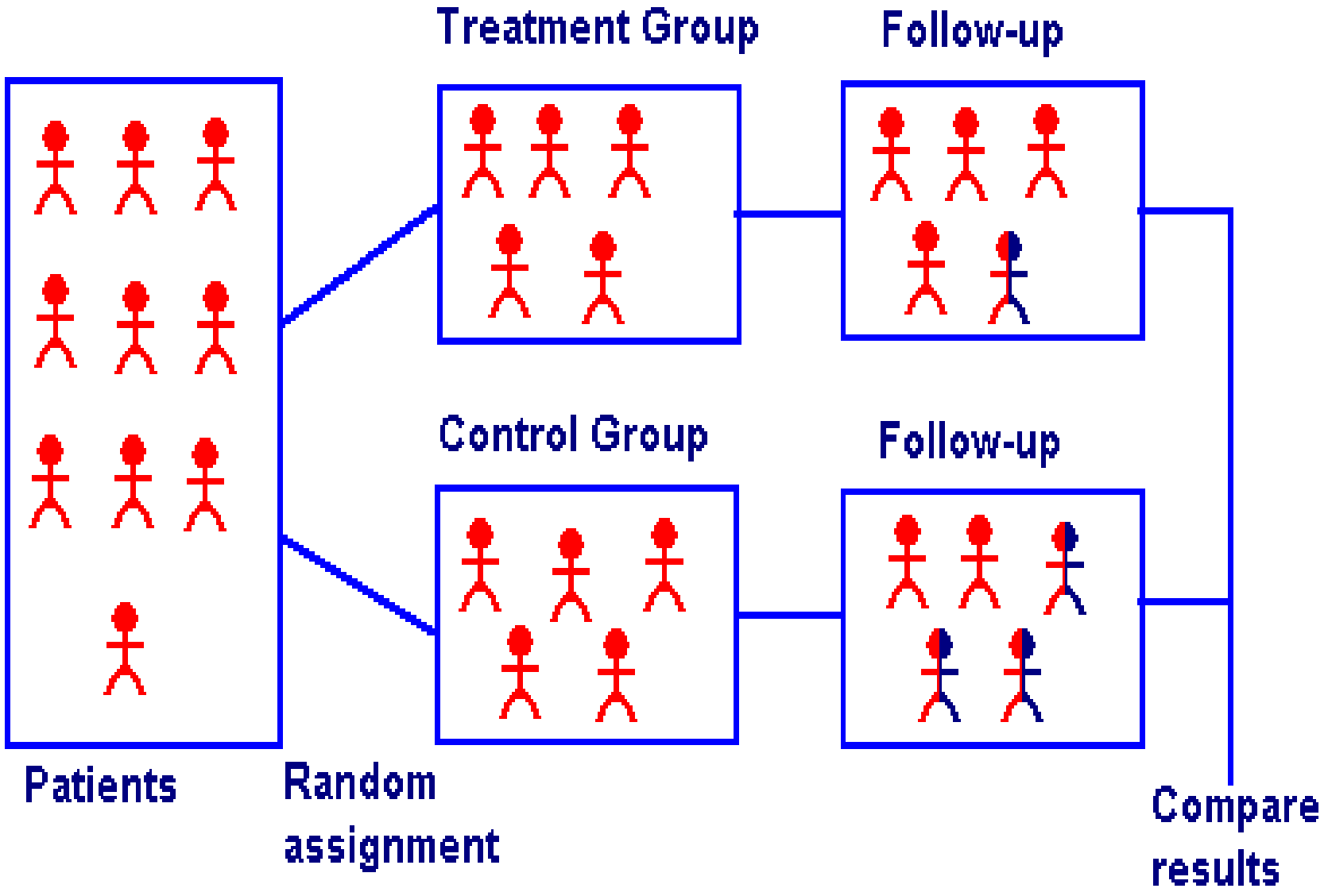
Clinical Trial Compares

- **INTERVENTION**
  - Drug (New)
  - Structured exercise program (e.g. osteoporosis)
  - Surgical procedure

- **CONTROL**
  - Placebo, old drug or old intervention
  - Usual regular advise given (osteoporosis)
  - Another surgical procedure / No surgery

Process of RCTs

- ✓ Preparation: Randomization, Computer generated list
- ✓ Eligibility assessment (Inclusion/exclusion)
- ✓ Consent
- ✓ Allocation to study arms (Concealment)
- ✓ Baseline assessment
- ✓ Initiation of intervention (Blind)
- ✓ Follow-up
- ✓ Outcome assessment
- ✓ Data analysis


Treatment Group
Follow-up
Control Group
Follow-up
Patients
Random assignment
Compare results

Appraise the Evidence

- Assess **validity**? Correctness (likely to be true)
- What are the **results**? Clinically important
- Can we **apply** the results to our patient? Applicable in and useful for my patients

Users' Guides to the Medical Literature

A Manual for Evidence-Based Clinical Practice

The Evidence-Based
Medicine Working Group

Edited by
Gordon Guyatt, MD
Drummond Rennie, MD
Robert Hayward, MD (interactive guides)

AMA press

JAMA & ARCHIVES JOURNALS
American Medical Association

Validity

- Internal validity: How well was the study done? Do the results reflect the *truth*?

- External validity: can I apply these results to MY patients?

VALIDITY

- Randomization.
- Concealment.
- Blindness.
- Follow up complete.
- Intention to treat.
- Similar groups at start.
- Both groups treated equally.

Are the results of this single preventive or therapeutic trial valid?

Was the assignment of patients to treatments randomised? Was the randomisation list concealed?	
Was follow-up of patients sufficiently long and complete?	
Were all patients analysed in the groups to which they were randomised?	
Were patients and clinicians kept "blind" to treatment?	
Were the groups treated equally, apart from the experimental treatment?	
Were the groups similar at the start of the trial?	

Randomization

- Randomisation = similar groups at baseline
- Equal (50%) chance to be in either group
- How was it randomized?
- Was randomization concealed?
  - selection
  - allocation

Was allocation assignment "concealed"?

Did investigators know to which group the potential subject would be assigned **before** enrolling them?

Importance of concealed allocation

Trials with unconcealed allocation consistently overestimate benefit by ~40%

Schulz KF, Chalmers I, Hayes RJ, et al. JAMA 1995;273:408-12
Schulz KF, Grimes DA. Lancet 2002;359:614-18.
Pildal J, et al. Int J Epidemiol 2007;36:847-857
Moher D, et al. Lancet 1998;352:609-13.

Ensuring Allocation Concealment

BEST – most valid technique

- Central computer randomization

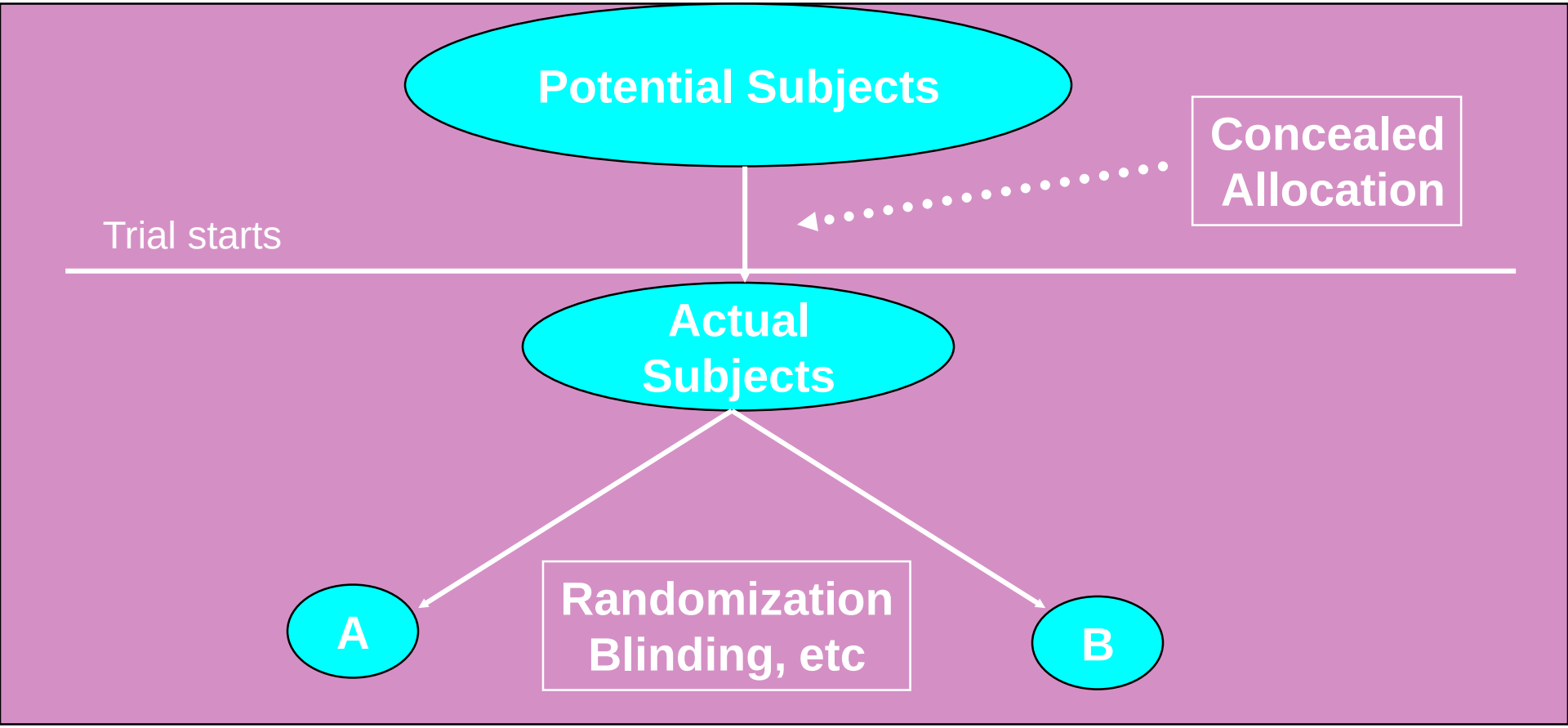
DOUBTFUL

- Envelopes, etc

NOT RANDOMIZED

Date of birth, alternate days, etc

Conducting a Study



Selection bias

Reduced by:

- ✓ centralised randomisation
- ✓ on-site computer system with group assignments in a locked file
- ✓ sequentially numbered, sealed, opaque envelopes

- ✗ Not: alternation, dates of birth, day of week.

Blindness

Was study "double-blinded"?

- Did the ***patients*** know to which group they were assigned?
- Did the ***treating*** physician know?
- Did ***investigators*** assessing outcomes know ("triple-blinding" – up to 7 levels!)?
  - Judicial assessor blind + allocation concealment = surgery RCTs

Schulz KF. Ann Int Med 2002;136:254-9.

Measurement Bias - minimizing differential

- Blinding – Who?
  - Participants?
  - Investigators?
  - Outcome assessors?
  - Analysts?

- Most important to use "blinded" outcome assessors when outcome is not objective!

- Papers should report WHO was blinded and HOW it was done

Figure 1: The authors: double blinded versus single blinded

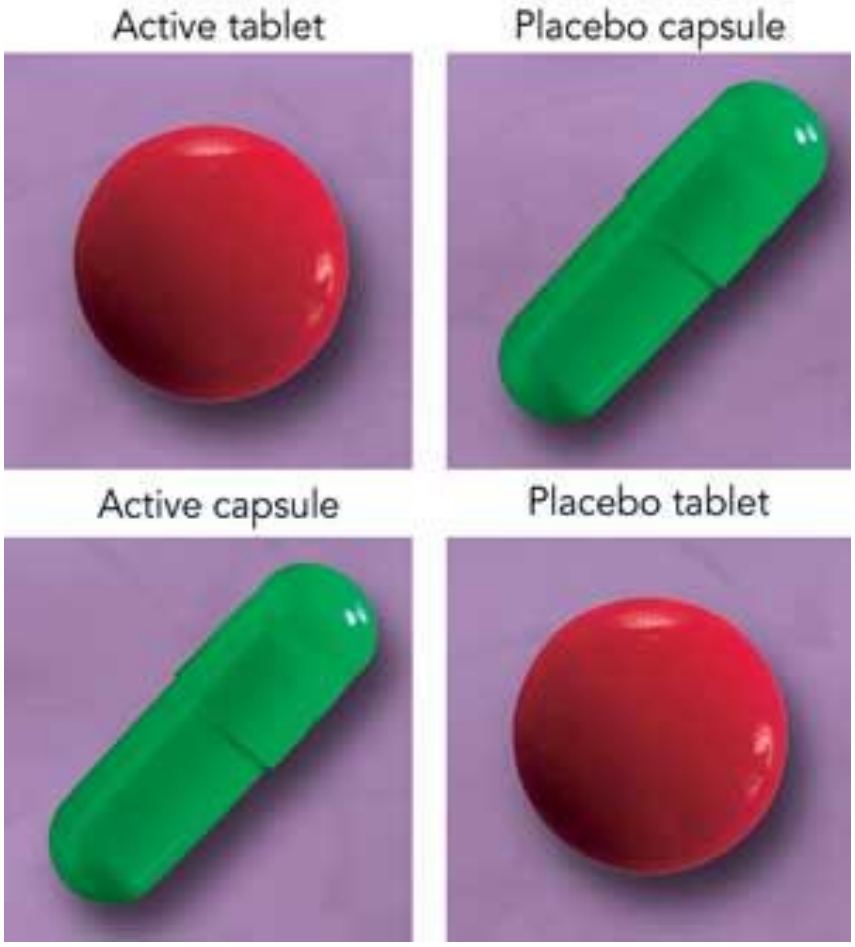
Figure 2: The authors blinded and masked

Schulz and Grimes. Lancet, 2002

Best RCTs: Double Blind

- Subject doesn't know which he's getting.
- Researcher doesn't know which he's giving.

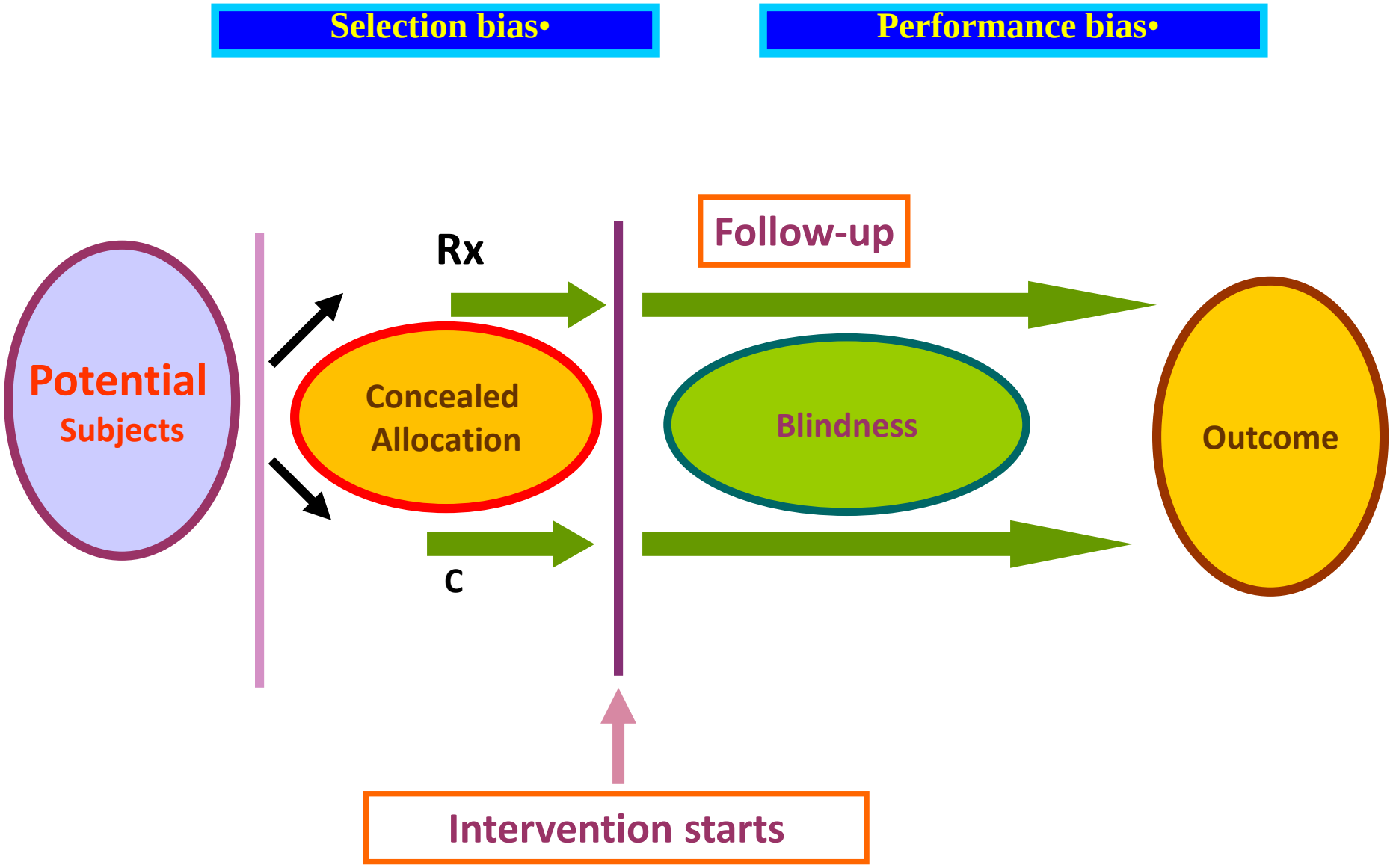
- Exit poll to see if patients could guess if they were in the placebo group


Active tablet
Placebo capsule
Active capsule
Placebo tablet

Blindness

- If patient knows: ***Placebo effect*** Those who are on effective treatment perform better than those who receive Placebo

- If Physician knows: ***Overestimate Treatment effect*** (More care, Co-intervention)


Selection bias•
Performance bias•
Potential
Subjects
Rx
C
Concealed
Allocation
Follow-up
Blindness
Outcome
Intervention starts

How RCTs differ from other designs

Two balanced groups:

- **Start Balanced:** All prognostic factors are equally distributed at the start (Concealed Randomization)
- **Run Balanced:** All prognostic factors are maintained balanced throughout the study (Blindness and the 3C)
- **End Balanced:** All prognostic factors are maintained balanced at the end of the study (ITT)

Intervention

ORIGINAL ARTICLE

Irbesartan in Patients with Heart Failure and Preserved Ejection Fraction

Barry M. Massie, M.D., Peter E. Carson, M.D., John J. McMurray, M.D., Michel Komajda, M.D., Robert McKelvie, M.D., Michael R. Zile, M.D., Susan Anderson, M.S., Mark Donovan, Ph.D., Erik Iverson, M.S., Christoph Staiger, M.D., and Agata Ptaszynska, M.D., for the I-PRESERVE Investigators*

ABSTRACT

BACKGROUND

Approximately 50% of patients with heart failure have a left ventricular ejection fraction of at least 45%, but no therapies have been shown to improve the outcome of these patients. Therefore, we studied the effects of irbesartan in patients with this syndrome.

METHODS

We enrolled 4128 patients who were at least 60 years of age and had New York Heart Association class II, III, or IV heart failure and an ejection fraction of at least 45% and randomly assigned them to receive 300 mg of irbesartan or placebo per day. The primary composite outcome was death from any cause or hospitalization for a cardiovascular cause (heart failure, myocardial infarction, unstable angina, arrhythmia, or stroke). Secondary outcomes included death from heart failure or hospitalization for heart failure, death from any cause and from cardiovascular causes, and quality of life.

RESULTS

During a mean follow-up of 49.5 months, the primary outcome occurred in 742 patients in the irbesartan group and 763 in the placebo group. Primary event rates in the irbesartan and placebo groups were 100.4 and 105.4 per 1000 patient-years, respectively (hazard ratio, 0.95; 95% confidence interval [CI], 0.86 to 1.05; P=0.35). Overall rates of death were 52.6 and 52.3 per 1000 patient-years, respectively (hazard ratio, 1.00; 95% CI, 0.88 to 1.14; P=0.98). Rates of hospitalization for cardiovascular causes that contributed to the primary outcome were 70.6 and 74.3 per 1000 patient-years, respectively (hazard ratio, 0.95; 95% CI, 0.85 to 1.08; P=0.44). There were no significant differences in the other prespecified outcomes.

CONCLUSIONS

Irbesartan did not improve the outcomes of patients with heart failure and a preserved left ventricular ejection fraction. (ClinicalTrials.gov number, NCT00095238.)

From the University of California, San Francisco, and San Francisco Veterans Affairs Medical Center, San Francisco (B.M.M.); Georgetown University and Washington DC Veterans Affairs Medical Center, Washington, DC (P.E.C.); British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, United Kingdom (J.J.M.); Université Paris 6 and Hospital Pitié–Salpêtrière, Paris (M.K.); Hamilton Health Sciences, McMaster University, Hamilton, ON, Canada (R.M.); Ralph H. Johnson Veterans Affairs Medical Center and Medical University of South Carolina, Charleston (M.R.Z.); University of Wisconsin, Madison (S.A., E.I.); Bristol-Myers Squibb, Princeton, NJ (M.D., A.P.); and Sanofi-Aventis, Bridgewater, NJ (C.S.). Address reprint requests to Dr. Massie at the Veterans Affairs Medical Center, 111C, 4150 Clement St., San Francisco, CA 94121, or at barry.massie@va.gov.

*Committee members and investigators in the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE) are listed in the Supplementary Appendix, available with the full text of this article at www.nejm.org.

This article (10.1056/NEJMoa0805450) was published at www.nejm.org on November 11, 2008.

N Engl J Med 2008;359:2456-67.

Approximately half of patients with a diagnosis of heart failure have a normal or near-normal left ventricular ejection fraction.[1-5] Such patients differ from those with heart failure and a low left ventricular ejection fraction in a number of important ways: they tend to be older and female, and their condition is more likely to be associated with hypertension than with ischemia. The rates of death and illness among these patients are high and have not declined, as they have in patients with heart failure and a low left ventricular ejection fraction.[6]

Unfortunately, no pharmacologic therapy has been shown to be effective in improving outcomes in patients with heart failure with a preserved left ventricular ejection fraction. However, because the renin–angiotensin–aldosterone system is involved in many of the processes associated with this syndrome (including hypertension, left ventricular hypertrophy, myocardial fibrosis, and vascular dysfunction),[7,8] inhibitors of this system have been of particular interest as a therapeutic intervention for these patients.[9,10] Although information about neurohormone levels in this syndrome is limited, available data indicate that plasma renin activity is increased in patients with heart failure and a preserved left ventricular ejection fraction, as compared with control subjects, although levels are lower than in patients who have heart failure with a low left ventricular ejection fraction.[7] Furthermore, blockade of the renin–angiotensin system has had favorable effects in patients with a low left ventricular ejection fraction. It has also improved outcomes in patients after myocardial infarction, in those with hypertension, and in those with other high-risk vascular disease — populations that are thought to be at risk for heart failure with a preserved left ventricular ejection fraction.

Accordingly, we conducted the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE) to evaluate the effect of the angiotensin-receptor blocker irbesartan on mortality and cardiovascular morbidity in patients with heart failure and a preserved left ventricular ejection fraction.

METHODS

PATIENTS

We enrolled patients from centers in 25 countries. All patients were at least 60 years of age and had heart failure symptoms and a left ventricular ejection fraction of at least 45%.[11,12] In addition, we required patients to have been hospitalized for heart failure during the previous 6 months and have current New York Heart Association (NYHA) class II, III, or IV symptoms with corroborative evidence; if they had not been hospitalized, they were required to have ongoing class III or IV symptoms with corroborative evidence. Such evidence could include findings of pulmonary congestion on radiography, left ventricular hypertrophy or left atrial enlargement on echocardiography, or left ventricular hypertrophy or left bundle-branch block on electrocardiography. Treatment with an angiotensin-converting–enzyme (ACE) inhibitor was permitted only when such therapy was considered essential for an indication other than uncomplicated hypertension.

Exclusion criteria included previous intolerance to an angiotensin-receptor blocker; an alternative probable cause of the patient's symptoms (e.g., significant pulmonary disease); any previous left ventricular ejection fraction below 40%; a history of acute coronary syndrome, coronary revascularization, or stroke within the previous 3 months; substantial valvular abnormalities; hypertrophic or restrictive cardiomyopathy; pericardial disease; cor pulmonale or other cause of isolated right heart failure; a systolic blood pressure of less than 100 mm Hg or more than 160 mm Hg or a diastolic blood pressure of more than 95 mm Hg despite antihypertensive therapy; other systemic disease limiting life expectancy to less than 3 years; substantial laboratory abnormalities (such as a hemoglobin level of less than 11 g per deciliter, a creatinine level of more than 2.5 mg per deciliter [221 μmol per liter], or liver-function abnormalities); or characteristics that might interfere with compliance with the study protocol.

STUDY PROCEDURES

The trial was approved by the ethics committee at each participating center; all patients provided written informed consent. Eligible patients were treated with single-blind placebo for 1 to 2 weeks before randomization; those who successfully completed this run-in phase and whose condition remained clinically stable were randomly assigned in a 1:1 ratio to receive irbesartan or matching placebo. The randomization schedule was implemented with the use of an interactive voice-response system. The randomization block size was two and was stratified according to site. Patients were also stratified according to their use of an ACE inhibitor at randomization. Therefore, for each site, separate blocks of two were designated for patients who were taking an ACE inhibitor and for those who were not taking an ACE inhibitor. Randomization of patients who were taking an ACE inhibitor at baseline was capped at 33% at each site.

PATIENTS BLINDED

RANDOMIZATION

CONCEALED ALLOCATION

Patients were started on 75 mg of irbesartan or placebo once daily. The dose was doubled to 150 mg after 1 to 2 weeks and was doubled again to 300 mg after an additional 1 to 2 weeks, according to a forced-titration protocol as tolerated. In addition to the titration visits, patients were seen 8 weeks, 14 weeks, and 6 months after randomization and every 4 months thereafter. The score on the Minnesota Living with Heart Failure scale[13] and the plasma level of N-terminal pro-B–type natriuretic peptide (NT-proBNP) were recorded at randomization, at 6 and 14 months, and at the final study visit. Serum creatinine and potassium were measured before randomization and at weeks 2 and 8, at month 6, and annually thereafter and, along with NT-proBNP, were analyzed in a central laboratory (Esoterix Belgium).

TREATED THE SAME

The executive committee designed and oversaw the trial in collaboration with representatives of the study sponsors (Bristol-Myers Squibb and Sanofi-Aventis), with assistance from an international steering committee. The sponsors or a contract research organization collected the trial data, which were then analyzed at the Statistical Data Analysis Center at the University of Wisconsin, Madison, independently of the sponsors and according to a predefined statistical analysis plan. All investigators and committee members who were involved in the conduct of the study (except for members of the data and safety monitoring board) were unaware of study-group assignments. The manuscript was prepared and submitted for publication by members of the executive committee, who had unrestricted access to the study data and who vouch for the accuracy and completeness of the reported analyses.

DATA ANALYSTS BLINDED

INVESTIGATORS BLINDED

STUDY OUTCOMES AND DEFINITIONS

The primary outcome, which was analyzed as the time from randomization to the first event, was a composite of death from any cause or hospitalization for a protocol-specified cardiovascular cause. Reasons for such hospitalizations included worsening heart failure, myocardial infarction, stroke, unstable angina, ventricular or atrial dysrhythmia, or myocardial infarction or stroke that occurred during any hospitalization. The secondary outcomes were the components of the primary outcome (death from any cause and hospitalization for cardiovascular causes), a composite heart failure outcome (death due to worsening heart failure or sudden death or hospitalization due to worsening heart failure), a change in the total score on the Minnesota Living with Heart Failure scale at 6 months, a change in the plasma level of NT-proBNP at 6 months, a composite vascular-event outcome (death from cardiovascular causes, nonfatal myocardial infarction, or nonfatal stroke), and death from cardiovascular causes. Deaths and hospitalizations were adjudicated by members of an independent end-point committee who were unaware of study-group assignments and used prespecified criteria.

ADJUDICATORS BLINDED

STATISTICAL ANALYSIS

We originally anticipated an annual event rate of 18% for the primary outcome in the placebo group. A sample size of 3600 patients was planned to provide 1440 primary events, yielding a statistical power of 90% to detect a 14.5% reduction in risk with irbesartan, corresponding to a reduction in hazard of 15.75%, with a two-sided alpha of 0.05, assuming a recruitment period of 2 years and a minimum follow-up period of 2 years. A blinded review of event rates in 2004 indicated that outcomes had accumulated at a slower-than-anticipated rate. Consequently, to achieve the target number of events for the same decrease in the hazard in a reasonable time period, the sample size was increased to 4100 patients.

Data from all patients who underwent randomization were analyzed according to the intention-to-treat principle. The analyses of the primary outcome and other composites of death or hospitalization were performed with the use of Kaplan–Meier estimates, with the log-rank test for the comparison of the study groups, and a supportive Cox proportional-hazards model to calculate hazard ratios and 95% confidence intervals. Consistency of effects was assessed for eight prespecified subgroups, according to age (<65, 65 to 75, and >75 years), sex, ejection fraction (≤59% or >59%), the use or nonuse of ACE inhibitors and beta-blockers, the presence or ab-

INTENTION TO TREAT

sence of diabetes, hospitalization for heart failure within the previous 6 months, and geographic region (Europe, North America, or all other countries). Interactions were evaluated by fitting an interaction term between treatment and each of the eight covariates and then assessing significance with the use of a Wald test. The score on the Minnesota Living with Heart Failure scale and the log-transformed plasma level of NT-proBNP were studied by analysis of covariance, with the baseline value as a covariate. All analyses included the use of ACE inhibitors as a term in the model. To control for the global type I error, the study outcomes were examined in a prespecified sequence as described previously. If at any step superiority was not demonstrated at the 0.05 level, no conclusion would be drawn for subsequent outcomes. All P values are two-sided and were not adjusted for multiple testing.

The protocol specified that the data and safety monitoring board should conduct a single interim efficacy analysis for mortality from any cause after 50% of the total expected deaths had occurred. For this analysis, the Pocock approach was applied for harm and the O'Brien–Fleming approach was applied for benefit.

RESULTS

PATIENTS

From June 2002 through April 2005, a total of 4563 patients were formally screened and 4128 underwent randomization at 293 sites in 25 countries in Western Europe, Eastern Europe, North America, South America, South Africa, and Australia. Of those patients, 2067 were assigned to receive irbesartan and 2061 to receive placebo. The common study termination date was set for April 17, 2008, when it was estimated that at least 1440 events of the primary outcome would have occurred. The mean follow-up time was 49.5 months, and the trial included 16,798 patient-years of follow-up.

FOLLOW-UP

The study groups did not differ significantly in baseline characteristics (Table 1). The mean age was 72 years, and 60% of the patients were women. The primary cause of heart failure was hypertension in 64% of the patients and ischemic heart disease in 25%, and hypertension was present in 88% overall. Atrial fibrillation was present in 29% and diabetes mellitus in 27%. Forty-one percent of the patients were obese, which was defined as a body-mass index (the weight in kilograms divided by the square of the height in meters) of more than 30. At baseline, the median level of NT-ProBNP was 339 pg per milliliter (interquartile range, 133 to 964). Baseline medications included diuretics (83%, including 52% who were taking a loop diuretic), beta-blockers (59%), calcium-channel blockers (40%), spironolactone (15%), and ACE inhibitors (25%).

STUDY-DRUG ADMINISTRATION AND FOLLOW-UP

At the end of the titration phase, 84% of the patients in the irbesartan group and 88% of those in the placebo group had reached the 300-mg dose (mean doses, 275 mg and 284 mg, respectively). The proportion of patients reaching the target dose did not differ according to the use of an ACE inhibitor. During the study, the proportion of patients receiving an ACE inhibitor rose from 25% in the two groups at baseline to 39% in the irbesartan group and 40% in the placebo group, the use of spironolactone rose from 15% in the two groups at baseline to 28% in the irbesartan group and 29% in the placebo group, and the use of beta-blockers rose from 59% in the irbesartan group and 58% in the placebo group to 73% in the two groups.

TREATED THE SAME

Between baseline and 6 months, blood pressure declined by a mean (±SD) of 3.8±18.0 mm Hg systolic and 2.1±10.5 mm Hg diastolic in the irbesartan group and by a mean of 0.2±17.6 mm Hg systolic and 0.2±10.4 mm Hg diastolic in the placebo group; the decreases in the two groups persisted for the duration of the trial. Among the surviving patients, the discontinuation rates in the irbesartan group and in the placebo group, respectively, were 13% and 12% at 1 year, 21% and 20% at 2 years, and 34% and 33% at the end of the trial.

At the end of the study, vital-status data were not available for 29 patients (1%) in the irbesartan group and 44 patients (2%) in the placebo group. If contact could not be made at end of study, data for these patients were censored from the analysis at the date they were last known to be alive.

FOLLOW-UP

PRIMARY OUTCOME

The primary composite outcome occurred in 742 patients (36%) in the irbesartan group and in 763 patients (37%) in the placebo group. There were 100.4 end-point events per 1000 patient-years in

BASELINE CHARACTERISTICS

Table 1. Baseline Characteristics of the Patients.*

Characteristic	Placebo (N=2061)	Irbesartan (N=2067)
Demographic		
Age		
Mean — yr	72±7	72±7
≥75 yr — no. (%)	716 (35)	697 (34)
Female sex — no. (%)	1264 (61)	1227 (59)
Race — no. (%)†		
White	1925 (93)	1934 (94)
Black	43 (2)	39 (2)
Asian	15 (1)	19 (1)
Other	78 (4)	75 (4)
Clinical		
NYHA class — no. (%)‡		
II	445 (22)	426 (21)
III	1562 (76)	1582 (77)
IV	53 (3)	59 (3)
Heart rate — beats/min	71±10	72±11
Blood pressure — mm Hg		
Systolic	136±15	137±15
Diastolic	79±9	79±9
Body-mass index	29.6±5.3	29.7±5.3
Electrocardiographic findings — no. (%)		
Left ventricular hypertrophy	624 (30)	636 (31)
Left bundle-branch block	169 (8)	167 (8)
Atrial fibrillation or flutter	344 (17)	353 (17)
Ejection fraction	0.60±0.09	0.59±0.09
Cause of heart failure — no. (%)		
Ischemia	500 (24)	536 (26)
Hypertension	1304 (63)	1318 (64)
Hospitalization for heart failure within previous 6 mo — no. (%)	906 (44)	910 (44)
Medical history — no. (%)		
Hypertension	1816 (88)	1834 (89)
Angina symptoms§	824 (40)	828 (40)
Unstable angina	149 (7)	166 (8)
Myocardial infarction	482 (23)	487 (24)
PCI or CABG	267 (13)	281 (14)
Atrial fibrillation	603 (29)	606 (29)
Diabetes mellitus	564 (27)	570 (28)
Stroke or transient ischemic attack	201 (10)	198 (10)

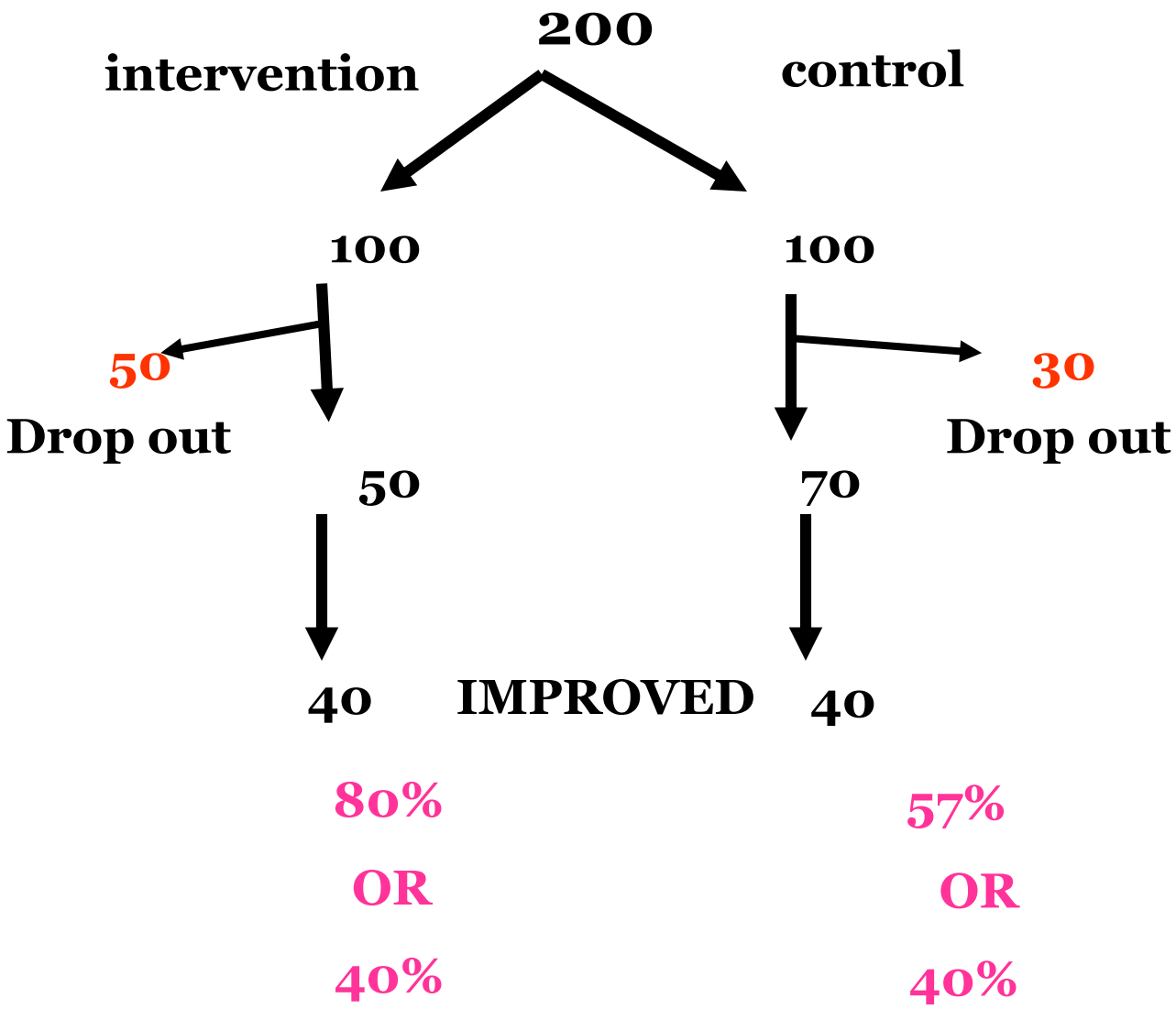
VALIDITY

- Randomization.
- Concealment.
- Blindness.
- Follow up complete.
- Similar groups at start.
- Both groups treated equally.
- Intention to treat.

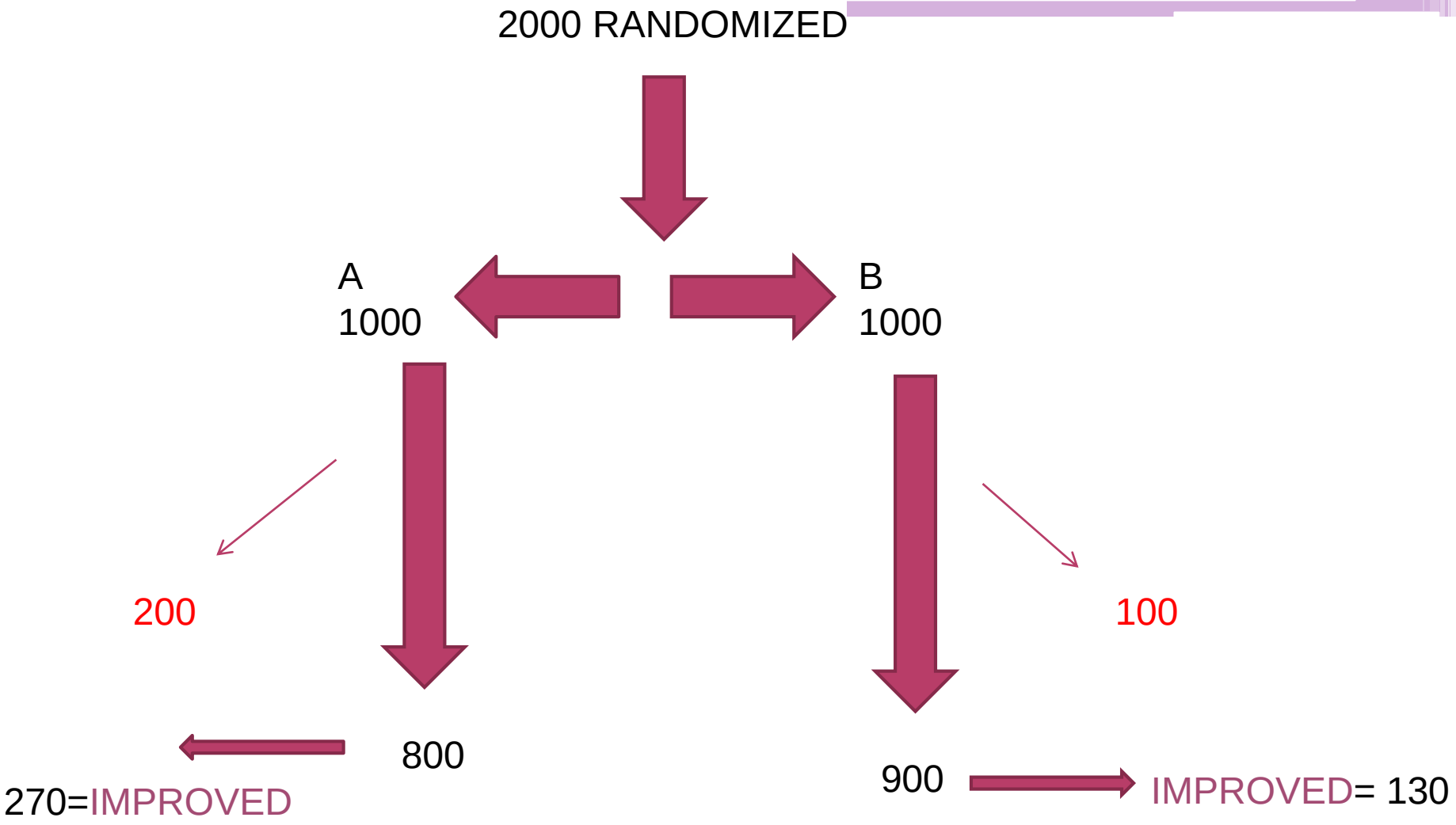
INTENTION TO TREAT

All patients analyzed in the groups to which they were allocated

INTENTION TO TREAT (ITT)

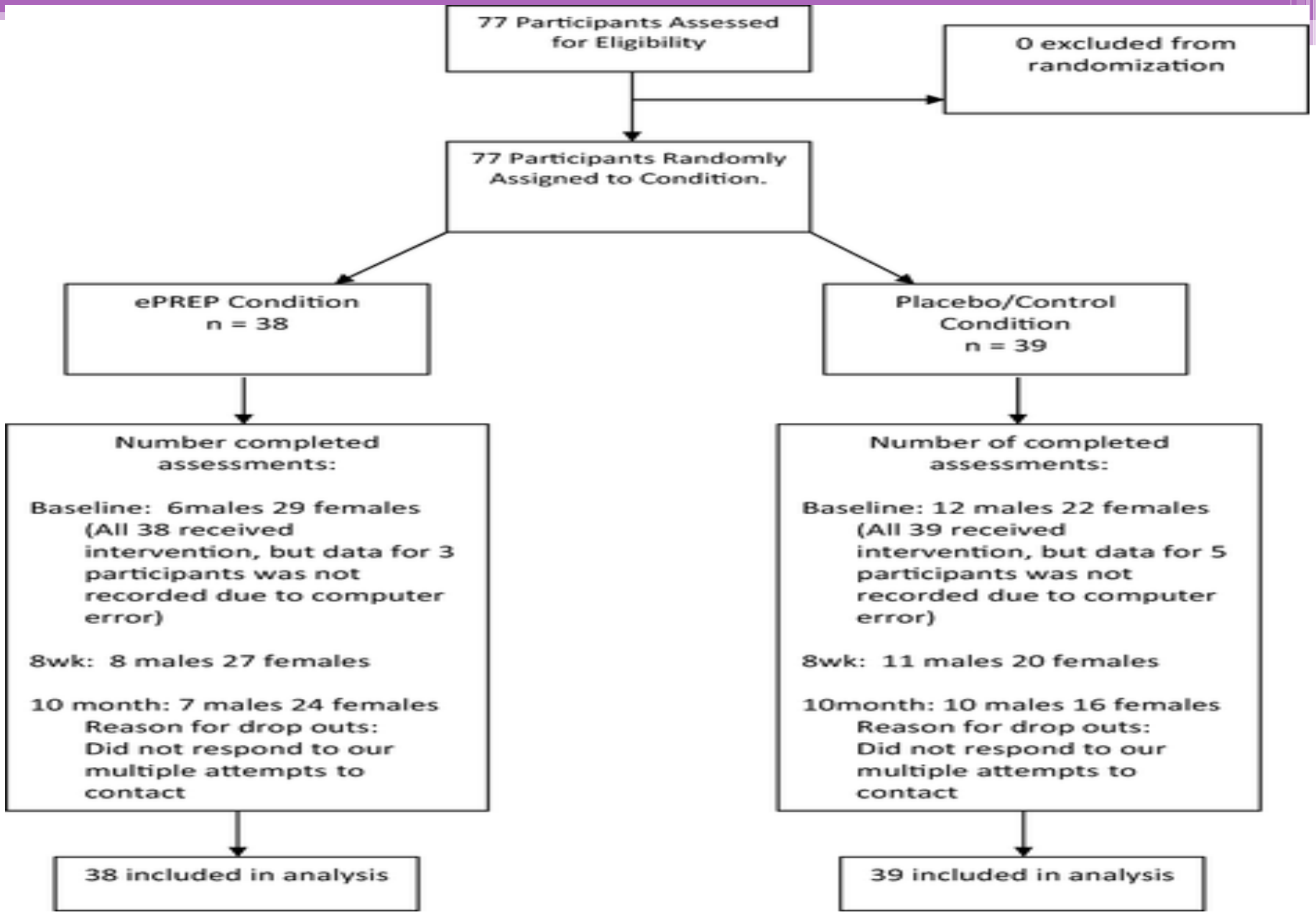


Were all patients analysed in the groups to which they were randomised?



EER= 270/?

CER=130/?


77 Participants Assessed for Eligibility
0 excluded from randomization
77 Participants Randomly Assigned to Condition.
ePREP Condition n = 38
Placebo/Control Condition n = 39
Number completed assessments:
Baseline: 6males 29 females (All 38 received intervention, but data for 3 participants was not recorded due to computer error)
8wk: 8 males 27 females
10 month: 7 males 24 females
Reason for drop outs: Did not respond to our multiple attempts to contact
Number of completed assessments:
Baseline: 12 males 22 females (All 39 received intervention, but data for 5 participants was not recorded due to computer error)
8wk: 11 males 20 females
10month: 10 males 16 females
Reason for drop outs: Did not respond to our multiple attempts to contact
38 included in analysis
39 included in analysis

Follow up

- duration of study.
- drop out < 20%.

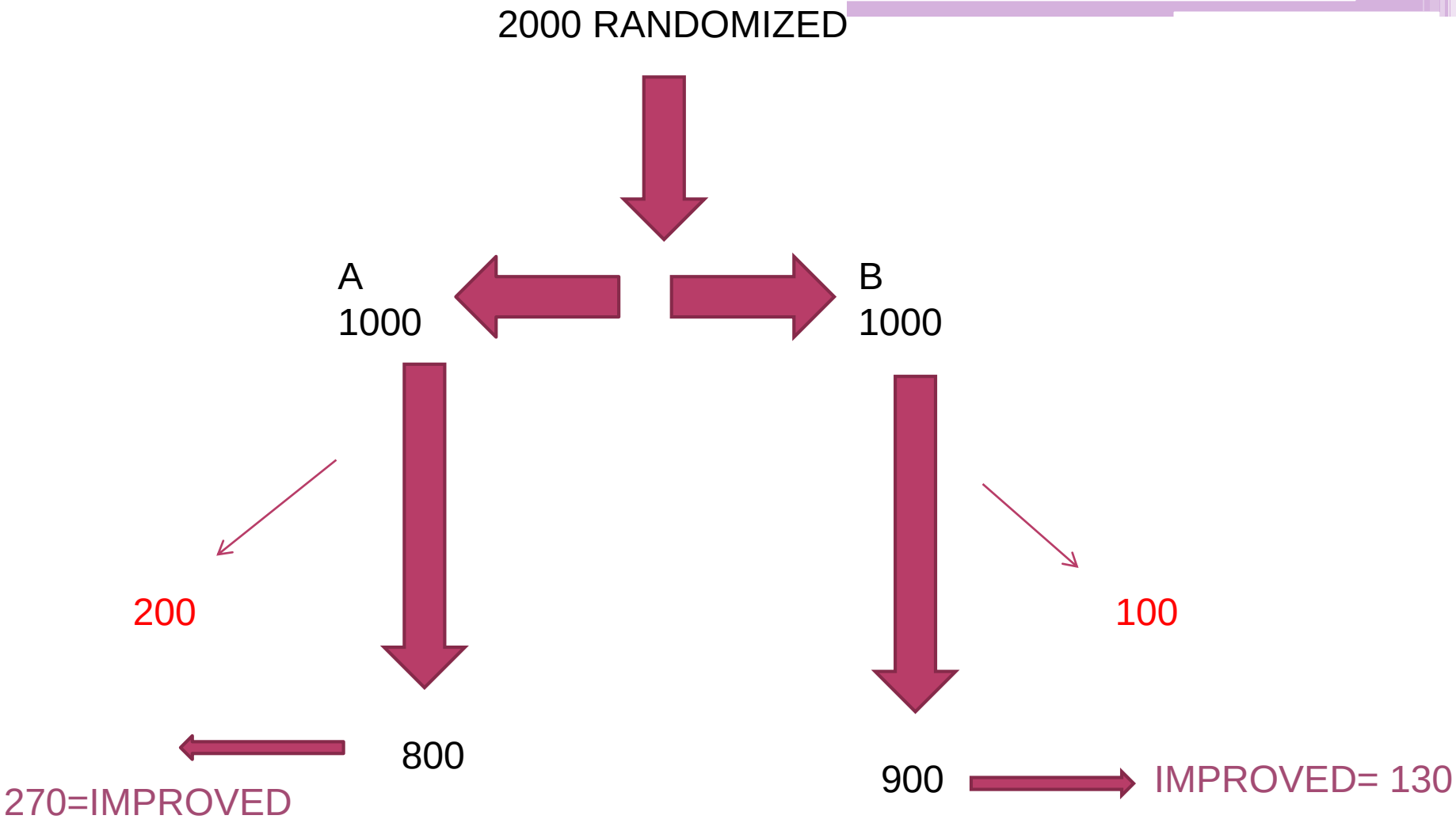
Was follow-up of patients sufficiently long and complete?

Results
“Importance”
Magnitude (treatment effect)

Definition

- Number Needed to Treat (NNT):
  - Number of persons who would have to receive an intervention for 1 to benefit.

NNT=1/ARR



EER= 270/?

CER=130/?

EER=270/800 = 33%= 0.33

CER= 130/900=14 %=0.14

ARR= 0.33-0.14= 0.19

NNT=1/0.19= 5.2=6

EER=270/1000=27%=0.27

CEER=130/1000= 13%=0.13

ARR=0.27-0.13=0.14

NNT=1/0.14=7

NUMBER NEED TO HARM(NNH)

WHAEN THE OUTCOME IS UNFAVOURABLE

Magnitude (treatment effect):

- Absolute effects (ARR & NNT)
- Relative effects (RR, RRR)

Result Tabulation

	Event + Ve	Event - Ve	Total
Experimental	a	b	a+b
Control	c	d	c+d

- EER = Experimental Event Rate (a/a+b)
- CER = Control Event Rate (c/c+d)

Result Tabulation

	Bleeding present	Bleeding Absent	Total
Drug A	20	80	100
Drug B	40	60	100

- EER-A (Risk A) = 20/100 = 20% (0.2)
- CER-B (Risk B) = 40/100 = 40% (0.4)

Calculations

- **ARR = CER - EER**
- **NNT = 1 / ARR**
- **RR = EER/CER (Risk A/Risk B)**
- **RRR = 1- RR**

Result Tabulation

	Bleeding present	Bleeding Absent	Total
Drug A	20	80	100
Drug B	40	60	100

- ARR = CER - EER NNT = 1 / ARR
- RR = EER/CER RRR = 1- RR

Calculations

- ARR = CER – EER = 0.4 – 0.2 = 0.2 (20%)
- NNT = 1 / ARR = 1/0.2 = 5
- RR = EER/CER = 0.2/0.4 = 0.5
- RRR = 1- RR = 1- 0.5= 0.5 (50%)

Precision

- **Confidence intervals:**
  - The range within which the likelihood of a true value is expected to be within a given degree of certainty, usually evaluated at 95% CI.
  - P value

APPLICABILITY

CAN I APPLY THESE VALID, IMPORTANT RESULTS TO MY PATIENT?

- **Do these results apply to my patient?**
  - **IS OUR PATIENT SO DIFFERENT?**
  - **IS THE TREATMENT FEASIBLE?**
  - **POTENTIAL BENEFITS AND HARMS**

- **Are my patient's values and preferences satisfied by the intervention offered?**

Summary

- Validity - is the paper likely to be true
- Importance - size of effect
  - NNT
  - Percision
- Applicability - can it work for me/my

DOES THIS STUDY ADDRESS A CLEAR QUESTION?

1. Were the following clearly stated:	Yes	Can't tell	No
• Patients • Intervention • Comparison Intervention • Outcome(s)			

ARE THE RESULTS OF THIS SINGLE TRIAL VALID?

A. The main questions to answer:

	Yes	Can't tell	No
2. Was the assignment of patients to treatments randomised? **3. Was the randomisation list concealed? Can you tell?**			
4. Were all subjects who entered the trial accounted for at it's conclusion? **5. Were they analysed in the groups to which they were randomised, i.e. intention-to-treat analysis**			

B. Some finer points to address:

	Yes	Can't tell	No
6. Were subjects and clinicians 'blind' to which treatment was being received, i.e. could they tell?			
7. Aside from the experimental treatment, were the groups treated equally?			
8. Were the groups similar at the start of the trial?			

WHAT WERE THE RESULTS?

	Outcome event		Total
	Yes	No	
Experimental group	a	b	a + b
Control group	c	d	c + d

Experimental event rate = risk of outcome event in experimental group = EER = a/(a+b)

Control event rate = risk of outcome event in control group = CER = c/(c+d)

Relative risk (RR) = $\frac{EER}{CER}$ Odds ratio (OR) = $\frac{ad}{bc}$

Relative risk reduction (RRR) = (CER - EER)/CER or 1 - RR

Absolute risk reduction (ARR) = CER - EER

Number needed to treat (NNT) = 1/ARR = 1/(CER - EER)

9. How large was the treatment effect? Consider • How were the results expressed (RRR, NNT, etc).	
10. How precise were the results? Were the results presented with confidence intervals?	

CAN I APPLY THESE VALID, IMPORTANT RESULTS TO MY PATIENT?

	Yes	Can't tell	No
11. Do these results apply to my patient? • Is my patient so different from those in the trial that the results don't apply? • How great would the benefit of therapy be for my particular patient?			
12. Are my patient's values and preferences satisfied by the intervention offered? • Do I have a clear assessment of my patient's values and preferences? • Are they met by this regimen and its potential consequences?			

Evaluating Research about Diagnostic Tests

Financial Disclosure: Dr Leischow reports consulting and serving as a paid speaker for Pfizer and as a consultant for Johnson & Johnson.

REFERENCES

1. Vevers V. Obama says WH will remain smoke-free. http://www.cbsnews.com/blogs/2008/12/08/politics/politicalhotsheet/entry4654231.shtml. Accessed December 9, 2008.
2. Reinberg S. Cancer to surpass heart disease as world's leading killer. washingtonpost.com Web page. December 9, 2008. http://www.washingtonpost.com/wp-dyn/content/article/2008/12/09/AR2008120901814.html. Accessed December 10, 2008.
3. Centers for Disease Control and Prevention. Cigarette smoking among adults—United States, 2007 [published correction appears in *MMWR Morb Mortal Wkly Rep*. 2008;57(47):1281]. *MMWR Morb Mortal Wkly Rep*. 2008; 57(45):1221-1226.
4. Centers for Disease Control and Prevention. Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000-2004. *MMWR Morb Mortal Wkly Rep*. 2008;57(45):1226-1228.
5. American Nonsmokers' Rights Foundation. Summary of 100% smokefree state laws and population protected by 100% US smokefree laws. January 4, 2009. http://www.no-smoke.org/pdf/SummaryUSPopList.pdf. Accessed December 12, 2009.
6. Risbeck CA. ADHA smoking cessation initiative liaisons, II: partnering with tobacco quitlines. November 2007. http://findarticles.com/p/articles/mi_m1ANQ/is_/ai_n25015057. Accessed December 9, 2008.
7. WHO Framework Convention on Tobacco Control: third session of the conference of the parties to the WHO FCTC. http://www.who.int/fctc/en/. Accessed January 10, 2009.
8. Fiore M, Croyle RT, Curry SJ, et al. Preventing 3 million premature deaths and helping 5 million smokers quit: a national action plan for tobacco cessation. *Am J Public Health*. 2004;94(2):205-210.
9. Peto R, Lopez A. The future worldwide health effects of current smoking patterns. In: Boyle P, Gray N, Henningfield J, Seffrin J, Zatonski W, eds. *Tobacco and Public Health: Science and Policy*. New York, NY: Oxford University Press; 2004: 281-286.
10. Mackay J, Erikson M, Shafey O. *The Tobacco Atlas*. 2nd ed. Atlanta, GA: American Cancer Society; 2006.
11. Fiore MC, Jaén CR, Baker TB, et al. Treating tobacco use and dependence: 2008 update. Rockville, MD: US Dept of Health and Human Services; May 2008. http://www.surgeongeneral.gov/tobacco/. Accessed February 5, 2009.
12. National Institute of Allergy and Infectious Dieseases. Treatment of HIV infection. http://www.niaid.nih.gov/factsheets/treat-hiv.htm. Updated November 8, 2007. Accessed February 8, 2009.
13. CEO Roundtable on Cancer Web page. http://www.ceoroundtableoncancer.org. Accessed December 19, 2008.

Diagnostic Errors—The Next Frontier for Patient Safety

David E. Newman-Toker, MD, PhD
Peter J. Pronovost, MD, PhD

DURING THE PAST DECADE, AWARENESS AND UNDERstanding of medical errors have expanded rapidly, with an energetic patient safety movement promoting safer health care through "systems" solutions. Efforts have focused on translating evidence into practice, mitigating hazards from therapies, and improving culture and communication. Diagnostic errors have received relatively little attention. Although the science of error measurement is underdeveloped, diagnostic errors are an important source of preventable harm.[1-3] In this Commentary, we offer definitions for diagnostic error and misdiagnosis-related harm, present an overview of the magnitude of diagnostic errors, and give suggestions for how research can mature.

Distinguishing Errors From Harms

In considering diagnostic errors, it is important to distinguish between the error (a process) and the resulting harm (an outcome). *Diagnostic error* can be defined as a diagnosis that is missed, wrong, or delayed, as detected by some subsequent definitive test or finding.[1] However, not all misdiagnoses result in harm, and harm may be due to either disease or intervention. *Misdiagnosis-related harm* can be defined as preventable harm that results from the delay or failure to treat a condition actually present (when the working diagnosis was wrong or unknown) or from treatment provided for a condition not actually present.

An estimated 40 000 to 80 000 US hospital deaths result from misdiagnosis annually.[4] Roughly 5% of autopsies reveal lethal diagnostic errors for which a correct diagnosis coupled with treatment could have averted death.[5] In the Harvard Medical Practice Study, physician errors resulting in adverse events were more likely to be diagnostic than drug-related (14% vs 9%), and misdiagnoses were more likely to be considered negligent (75% vs 53%) and to result in serious disability (47% vs 14%).[6] Not surprisingly, tort claims for diagnostic errors are nearly twice as common as claims for medication errors and result in the largest payouts.[7] As with all types of medical error, the human toll of misdiagnosis on an individual or family can be tremendous, particularly when a healthy patient experiences an adverse event.

Diagnostic errors often are unrecognized or unreported, and the science of measuring these errors (and their effects) is underdeveloped.[1,2] Available statistics consider neither deaths due to misdiagnosis in outpatients nor misdiagnosis-related morbidity and associated costs. For example, stroke, the leading cause of serious, long-term disability in the United States, affects 780 000 Americans annually.[8] Opportunities to prevent disabling stroke are missed when patients experiencing mild or transient warning symptoms receive misdiagnoses. According to a recent systematic review, 9% of all cerebrovascular events are missed initially, and the odds of misdiagnosis increase at least 5-fold when symptoms are mild or transient.[9]

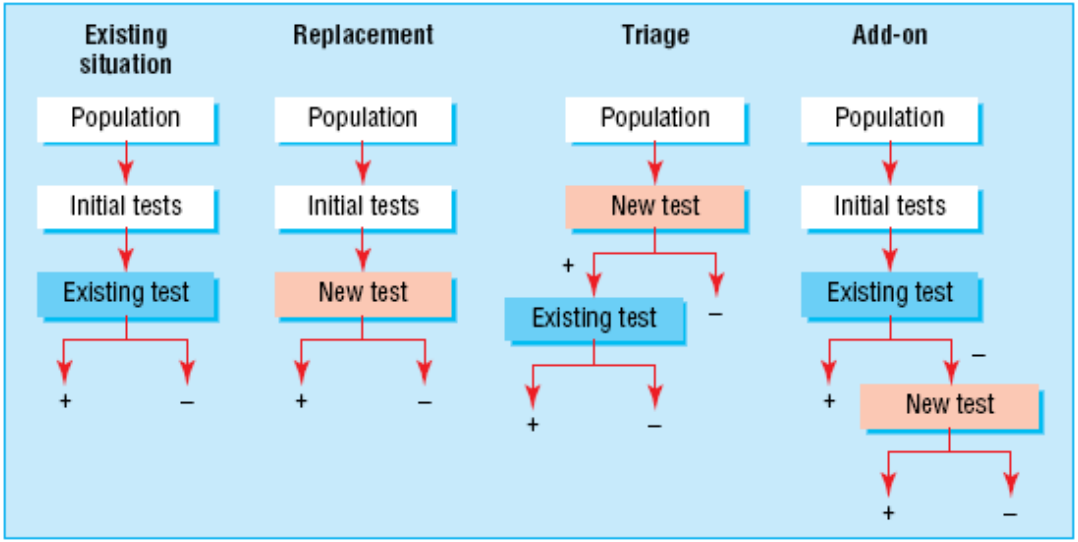
Author Affiliations: Departments of Neurology (Dr Newman-Toker) and Anesthesiology and Critical Care (Dr Pronovost), Johns Hopkins University School of Medicine, Baltimore, Maryland.
Corresponding Author: David E. Newman-Toker, MD, PhD, The Johns Hopkins Hospital, Pathology Bldg 2-210, 600 N Wolfe St, Baltimore, MD 21287 (toker@jhu.edu).

- 40,000-80,000 US hospital deaths from misdiagnosis per year
- Diagnosis uses <5% of hospital costs, but influences 60% of decision making

Roles of a new test

- Replacement – new replaces old
  - E.g., CT colonography for barium enema
- Triage – new determines need for old
  - E.g., B-natriuretic peptide for echocardiography
- Add-on – new combined with old
  - ECG and myocardial perfusion scan

Bossuyt et al BMJ 2006;332:1089–92



Roles of tests and positions in existing diagnostic pathways

Read this abstract

Patient-Initiated Treatment of Uncomplicated Recurrent Urinary Tract Infections in Young Women

Kalpana Gupta, MD, MPH; Thomas M. Hooton, MD; Pacita L. Roberts, MS; and Walter E. Stamm, MD

Background: Recurrent urinary tract infections (UTIs) are a common outpatient problem, resulting in frequent office visits and often requiring the use of prophylactic antimicrobial agents. Patient-initiated treatment of recurrent UTIs may decrease antimicrobial use and improve patient convenience.

Objective: To determine the safety and feasibility of patient-initiated treatment of recurrent UTIs.

Design: Uncontrolled, prospective clinical trial.

Setting: University-based primary health care clinic.

Participants: Women at least 18 years of age with a history of recurrent UTIs and no recent pregnancy, hypertension, diabetes, or renal disease.

Intervention: After self-diagnosing UTI on the basis of symptoms, participating women initiated therapy with ofloxacin or levofloxacin.

Measurements: Accuracy of self-diagnosis determined by evidence of a definite (culture-positive) or probable (sterile pyuria and no alternative diagnosis) UTI on pretherapy urinalysis and culture. Women with a self-diagnosis of UTI that was not microbiologically confirmed were evaluated for alternative diagnoses. Post-therapy interviews and urine cultures were used to assess clinical and microbiological cure rates, adverse events, and patient satisfaction.

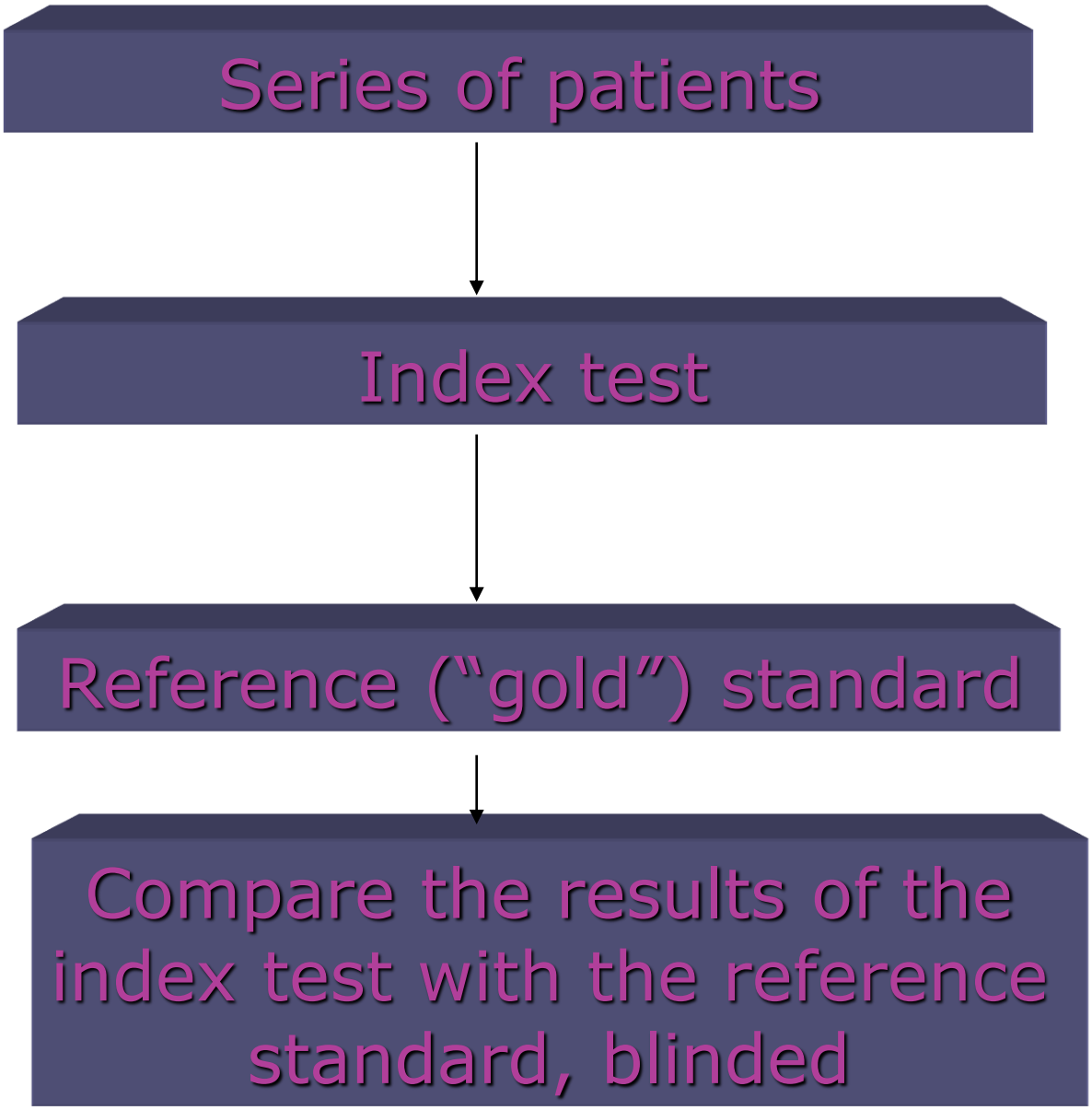
Results: 88 of 172 women self-diagnosed a total of 172 UTIs. Laboratory evaluation showed a uropathogen in 144 cases (84%), sterile pyuria in 19 cases (11%), and no pyuria or bacteriuria in 9 cases (5%). Clinical and microbiological cures occurred in 92% and 96%, respectively, of culture-confirmed episodes. No serious adverse events occurred.

Conclusion: Adherent women can accurately self-diagnose and self-treat recurrent UTIs.

Ann Intern Med. 2001;135:9-16. www.annals.org

For author affiliations, current addresses, and contributions, see end of text.

See related article on pp 41-50 and editorial comment on pp 51-52.


Series of patients
Index test
Reference (“gold”) standard
Compare the results of the index test with the reference standard, blinded

Patient-Initiated Treatment of Uncomplicated Recurrent Urinary Tract Infections in Young Women

Kalpana Gupta, MD, MPH; Thomas M. Hooton, MD; Pacita L. Roberts, MS; and Walter E. Stamm, MD

Background: Recurrent urinary tract infections (UTIs) are a common outpatient problem, resulting in frequent office visits and often requiring the use of prophylactic antimicrobial agents. Patient-initiated treatment of recurrent UTIs may decrease antimicrobial use and improve patient convenience.

Objective: To determine the safety and feasibility of patient-initiated treatment of recurrent UTIs.

Design: Uncontrolled, prospective clinical trial.

Setting: University-based primary health care clinic.

Participants: Women at least 18 years of age with a history of recurrent UTIs and no recent pregnancy, hypertension, diabetes, or renal disease.

Intervention: After self-diagnosing UTI on the basis of symptoms, participating women initiated therapy with ofloxacin or levofloxacin.

Measurements: Accuracy of self-diagnosis determined by evidence of a definite (culture-positive) or probable (sterile pyuria and no alternative diagnosis) UTI on pretherapy urinalysis and culture. Women with a self-diagnosis of UTI that was not microbiologically confirmed were evaluated for alternative diagnoses. Post-therapy interviews and urine cultures were used to assess clinical and microbiological cure rates, adverse events, and patient satisfaction.

Results: 88 of 172 women self-diagnosed a total of 172 UTIs. Laboratory evaluation showed a uropathogen in 144 cases (84%), sterile pyuria in 19 cases (11%), and no pyuria or bacteriuria in 9 cases (5%). Clinical and microbiological cures occurred in 92% and 96%, respectively, of culture-confirmed episodes. No serious adverse events occurred.

Conclusion: Adherent women can accurately self-diagnose and self-treat recurrent UTIs.

Ann Intern Med. 2001;135:9-16. www.annals.org

For author affiliations, current addresses, and contributions, see end of text.

See related article on pp 41-50 and editorial comment on pp 51-52.

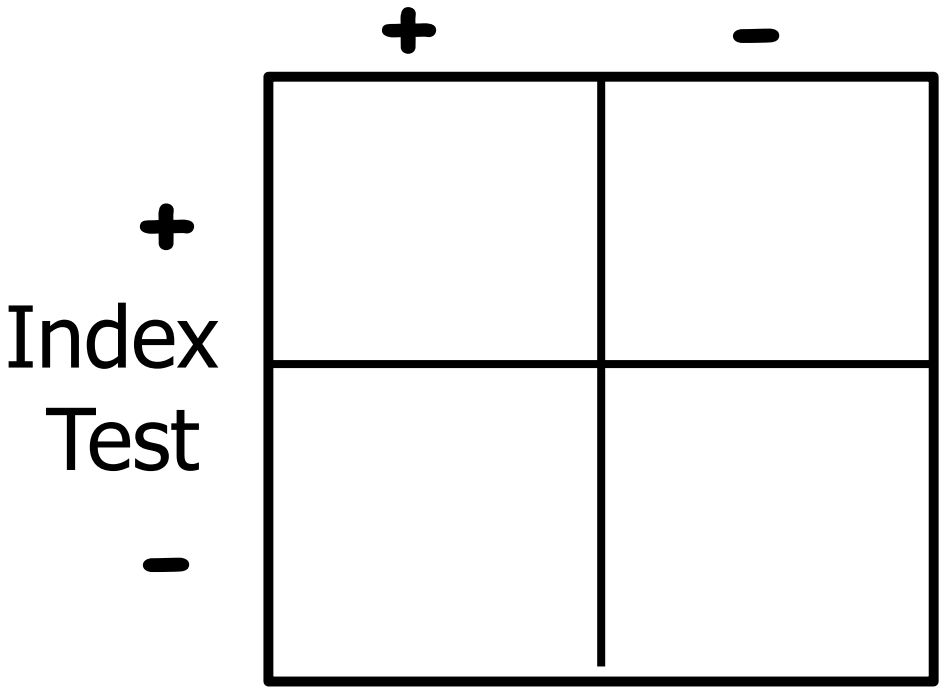
Reference standard

Accuracy

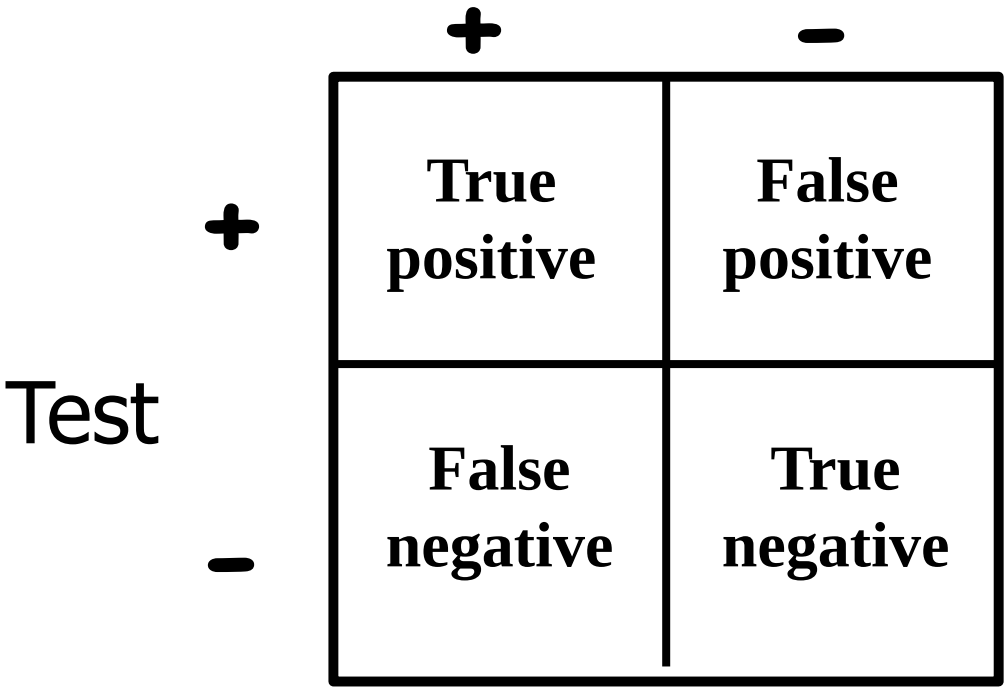
Series of patients

Index test

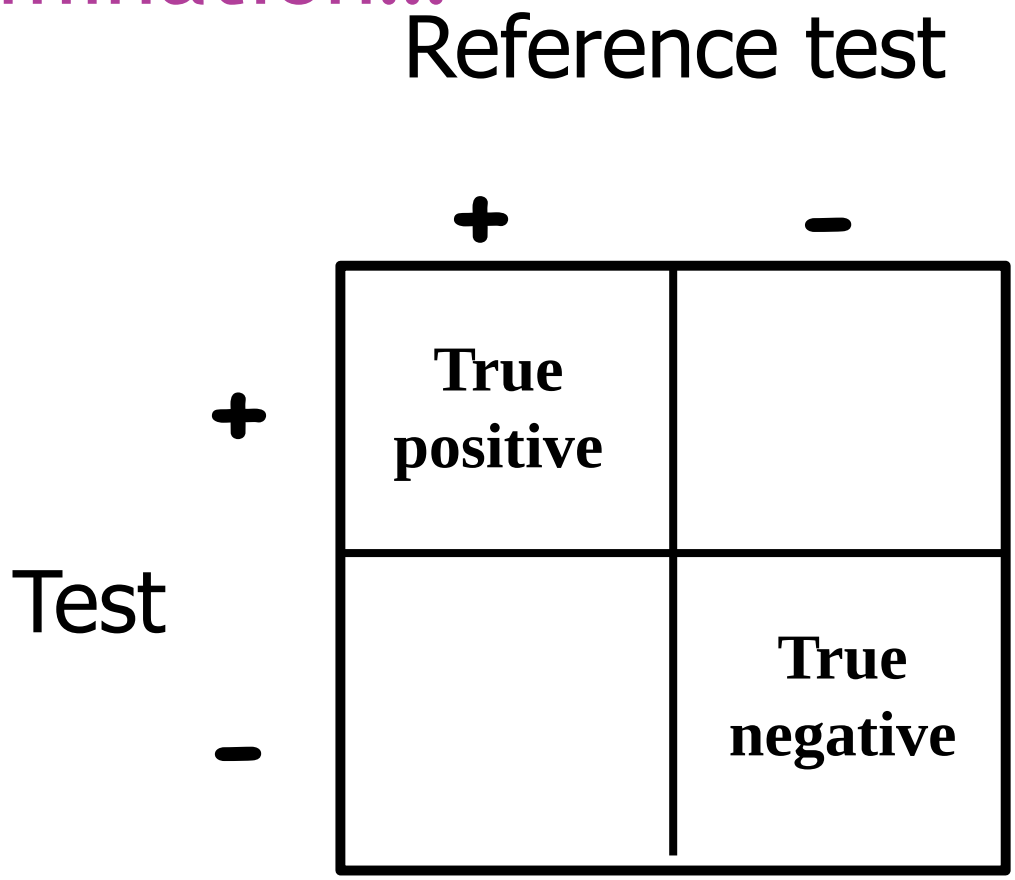
2 by 2 table



2 by 2 table


Reference test
+
-
Test
+
-
True positive
False positive
False negative
True negative

IF only a test had perfect discrimination…



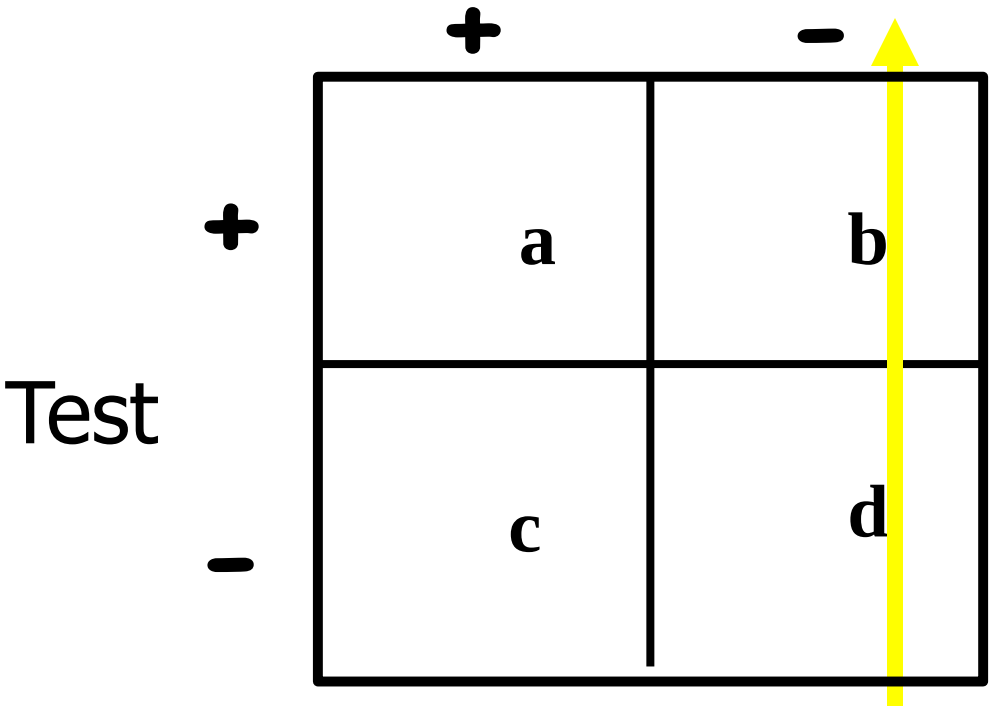
Sensitivity

Test \ Disease	+	−
+	a	b
−	c	d

Sensitivity = a / a + c

Proportion of people <u>with</u> the disease who have a <u>positive</u> test.

Specificity Disease



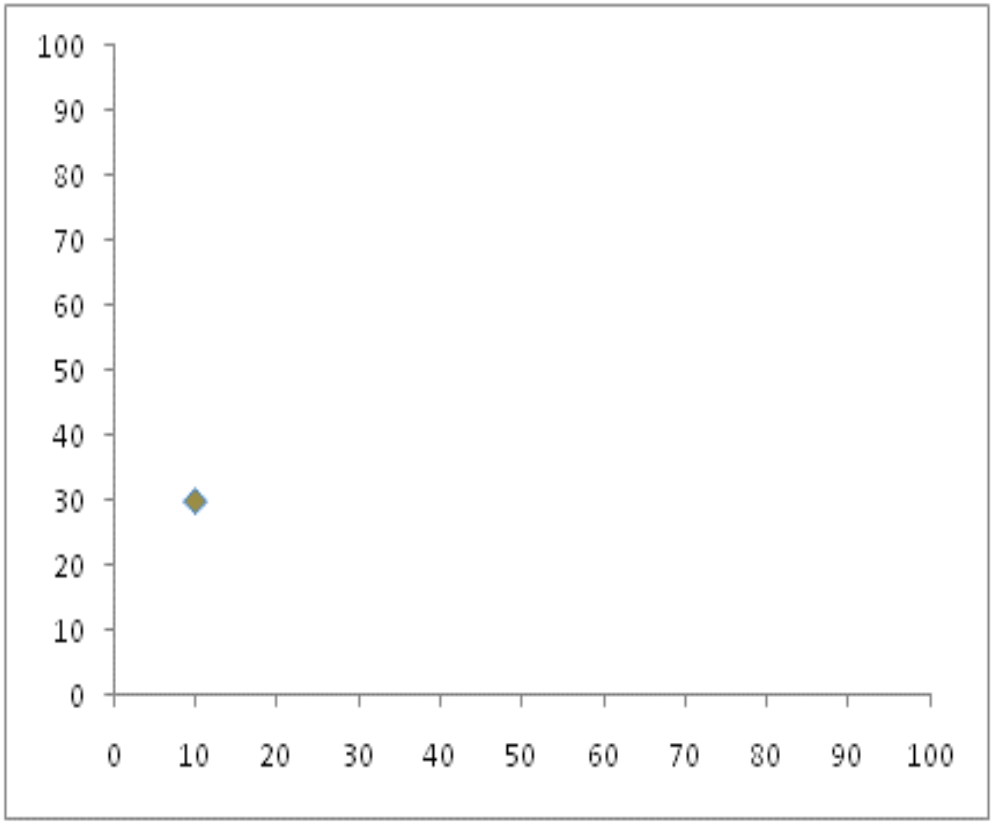
Proportion of people <u>without</u> the disease who have a negative test.

Specificity = d / b + d

ROC curves (Receiver Operating Characteristic curves) – What are they and what aren't they?

ROC curves – provide accuracy results over a *range of thresholds*

Sensitivity

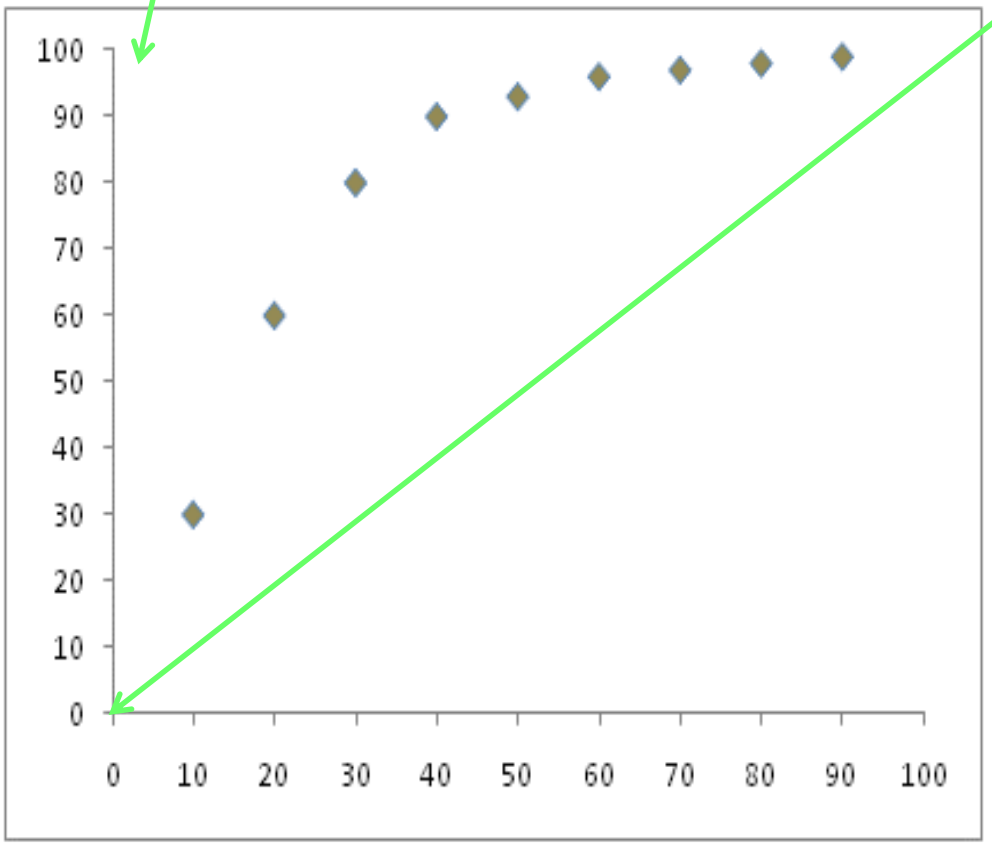


1-Specificity or false positive rate

A test with 30% sensitivity and 90% specificity (10% false positive rate) at one cut-point is plotted in the lower left corner.

Perfect test = upper left hand corner

Diagonal = no discrimination



Sensitivity

1-Specificity

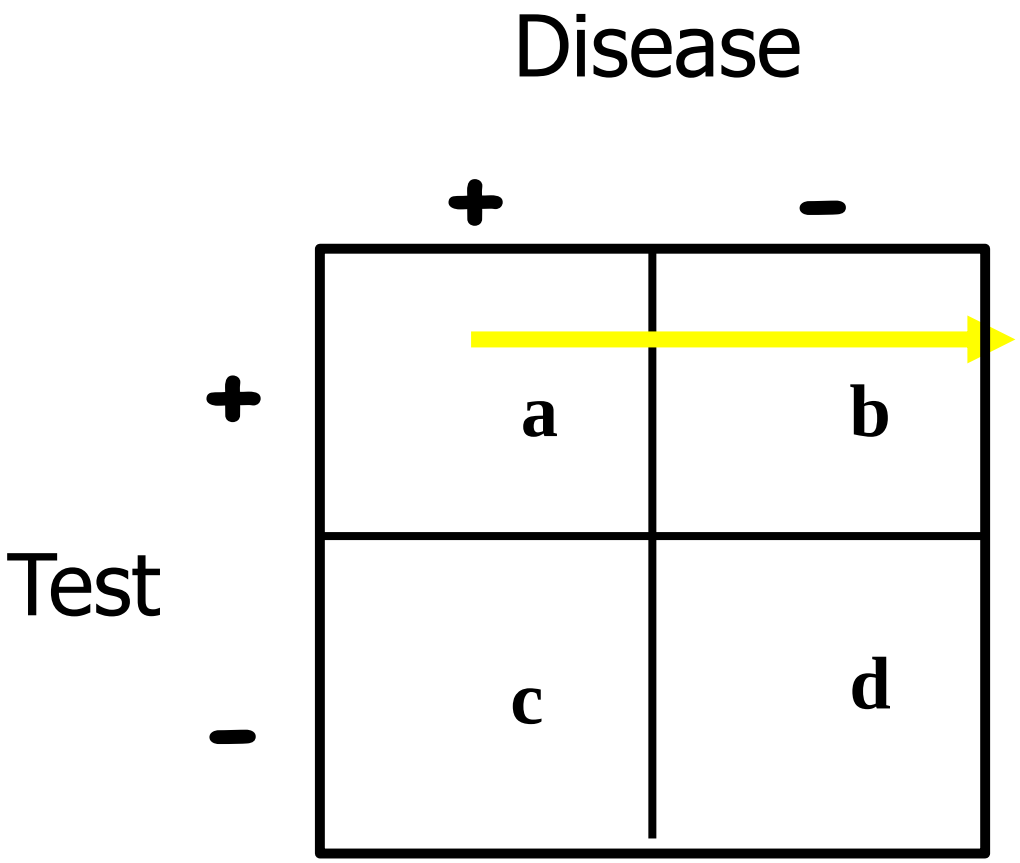
Area under the curve (AUC)
0.5 = useless
1.0 = perfect

Why this is important

http://today.msnbc.msn.com/id/42829175

What about positive and negative predictive values?

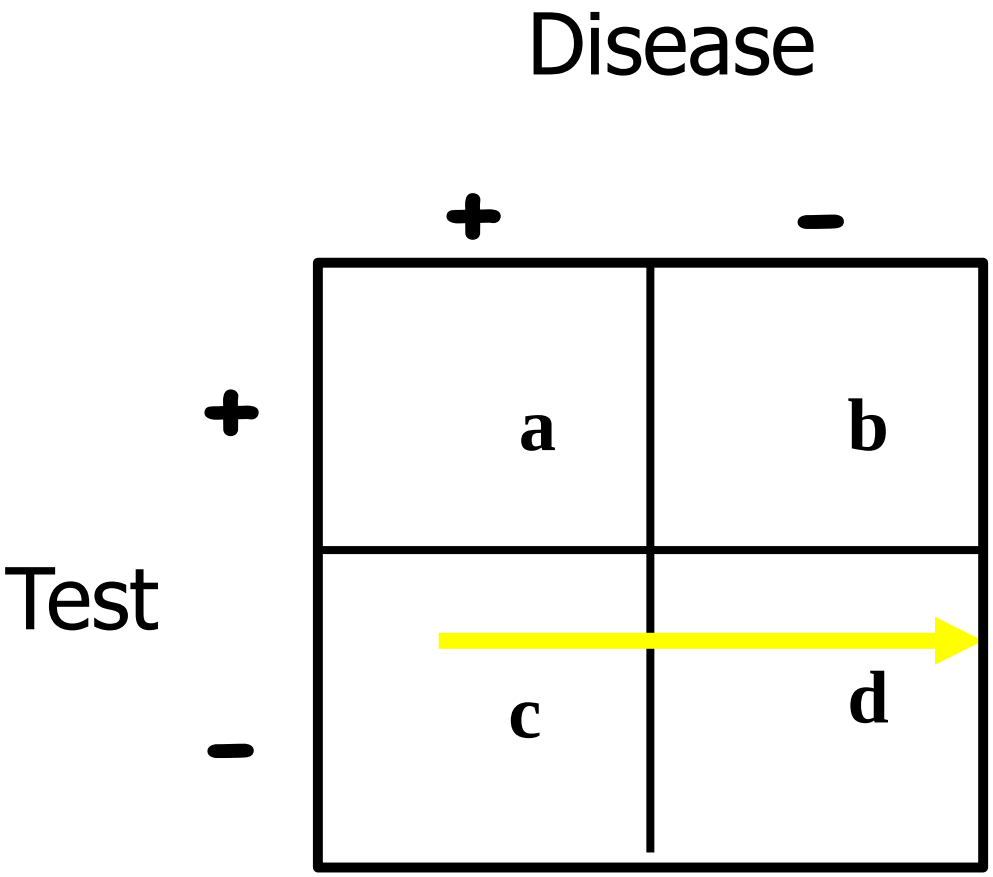
positive predictive value (PPV)



PPV = a / a + b

Proportion of people with a positive test who have the disease

negative predictive value (NPV)



NPV = d / c + d

Proportion of people with a <u>negative</u> test who <u>do not</u> have the disease

Sensitivity/specificity

- Disease status known
- Not as dependent on prevalence
- but can be affected by disease spectrum eg selection of patients

Positive/Negative predictive values

- Test result known
- Depend on prevalence

Likelihood Ratios and Bayesian reasoning

- Can use in situations with more than 2 test outcomes
- Direct link from pre-test probabilities to post-test probabilities

Positive and negative likelihood ratios

LR+ How much more often a positive test occurs in people <u>with</u> compared to those <u>without</u> the disease

LR+ = a/a+c / b/b+d

Or

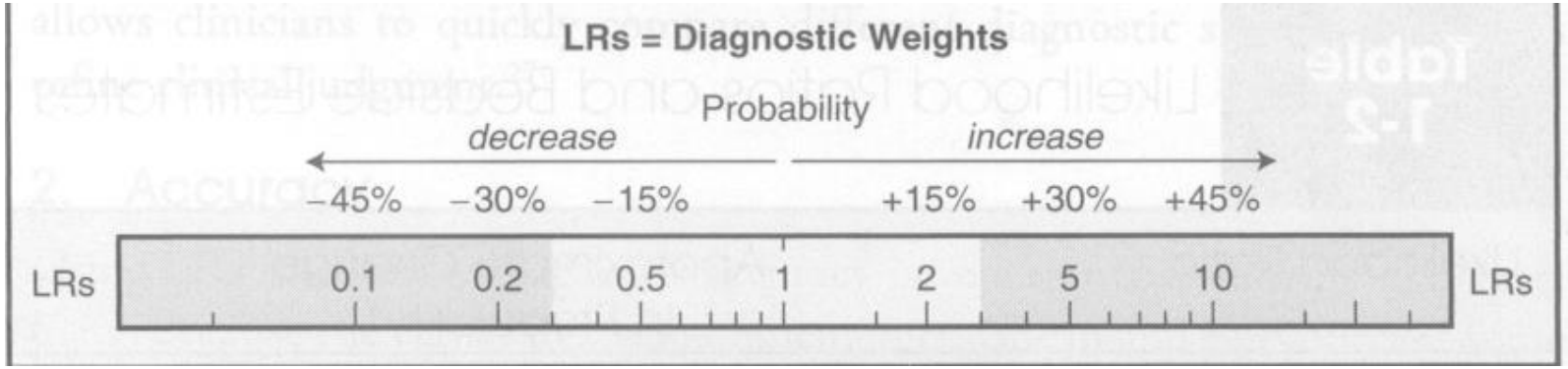
LR+ = sens/(1-spec)

LR- How less likely a negative test result is in people with the disease compared to those without the disease

LR- = c/a+c / d/b+d

Or

LR- = (1-sens)/(spec)

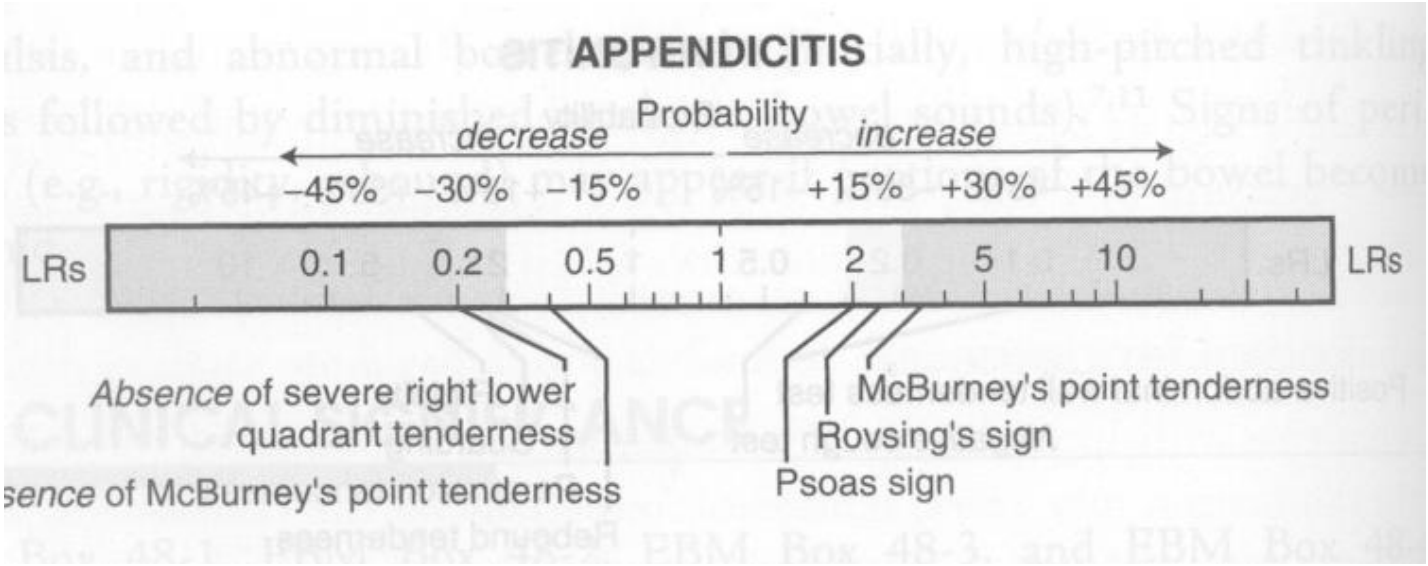


LR<0.1…. strong negative test result

LR=1

No diagnostic value

LR>10 …. strong positive test result



McGee: Evidence based Physical Diagnosis (Saunders Elsevier)

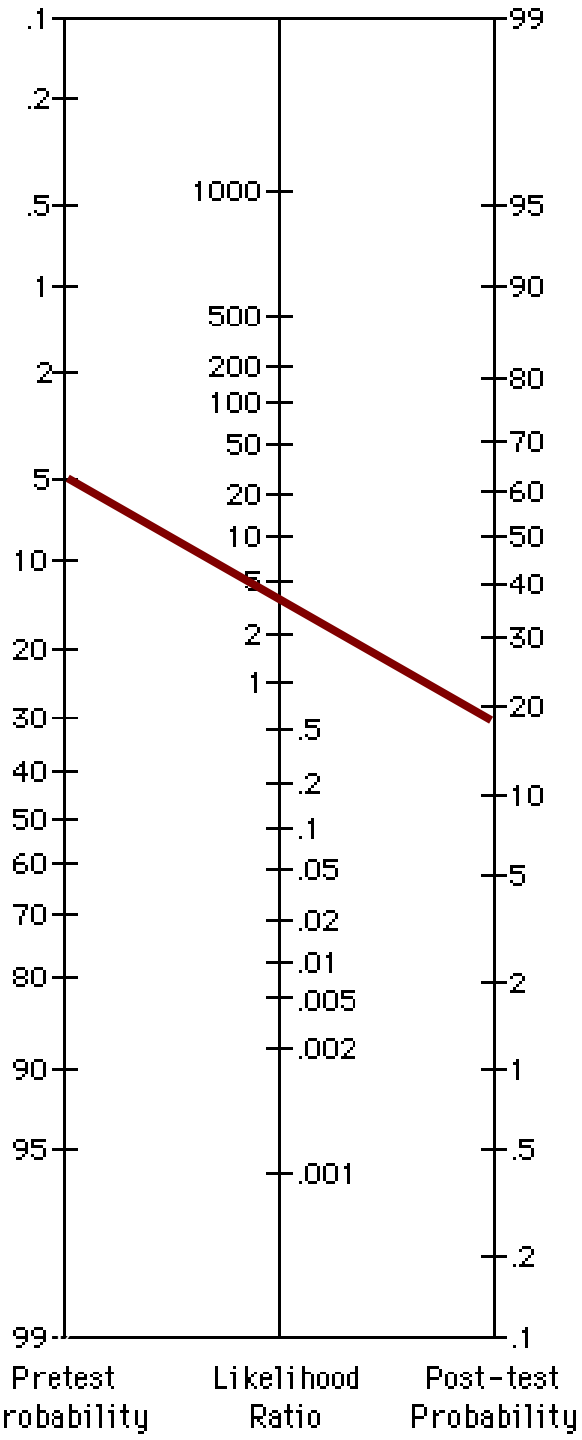
Bayesian reasoning

Post-test odds = Pre-test odds x Likelihood ratio

- Post-test odds for disease after *one* test become pre-test odds for *next* test etc

Bayesian reasoning using Fagan Nomogram

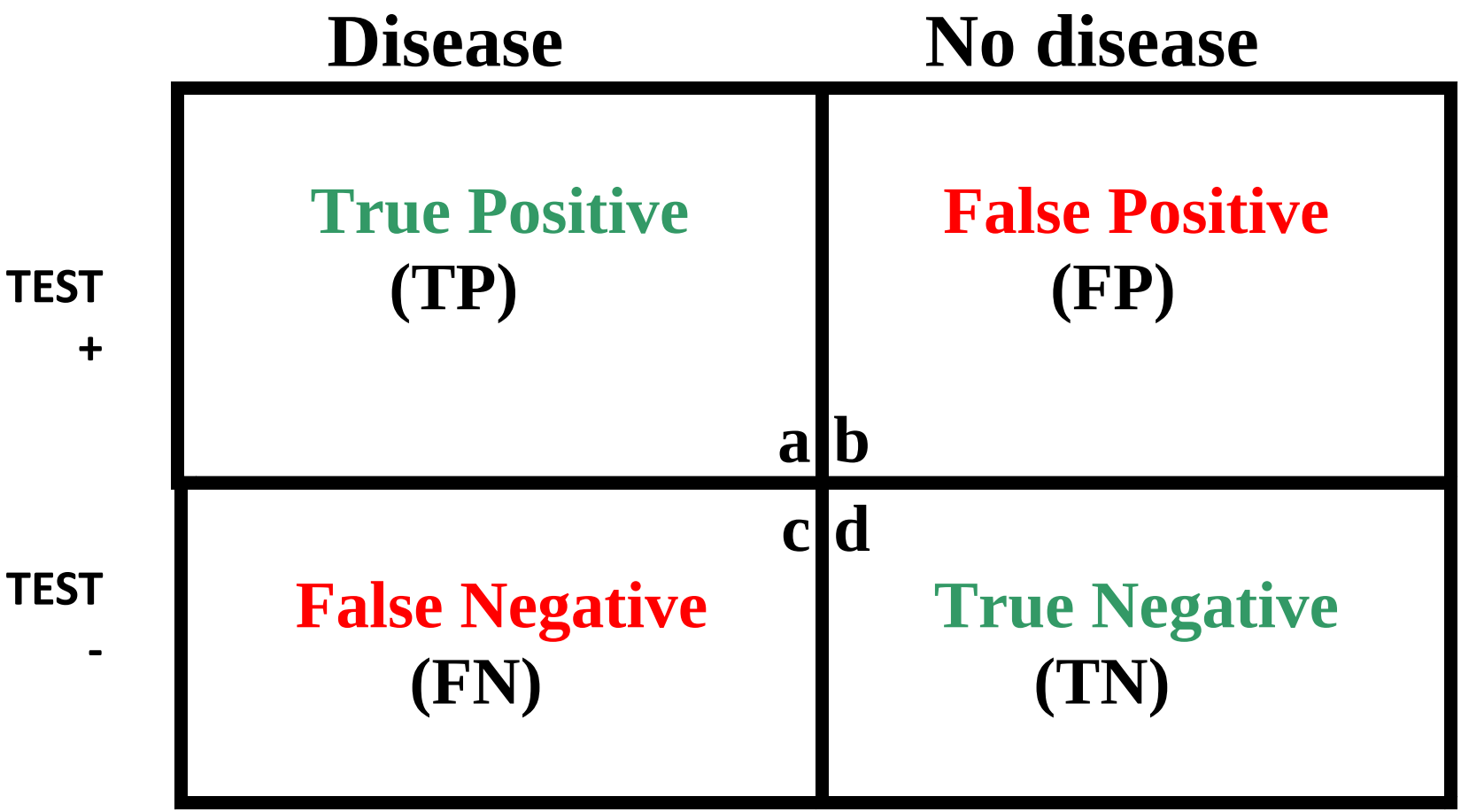
Pre test 5%



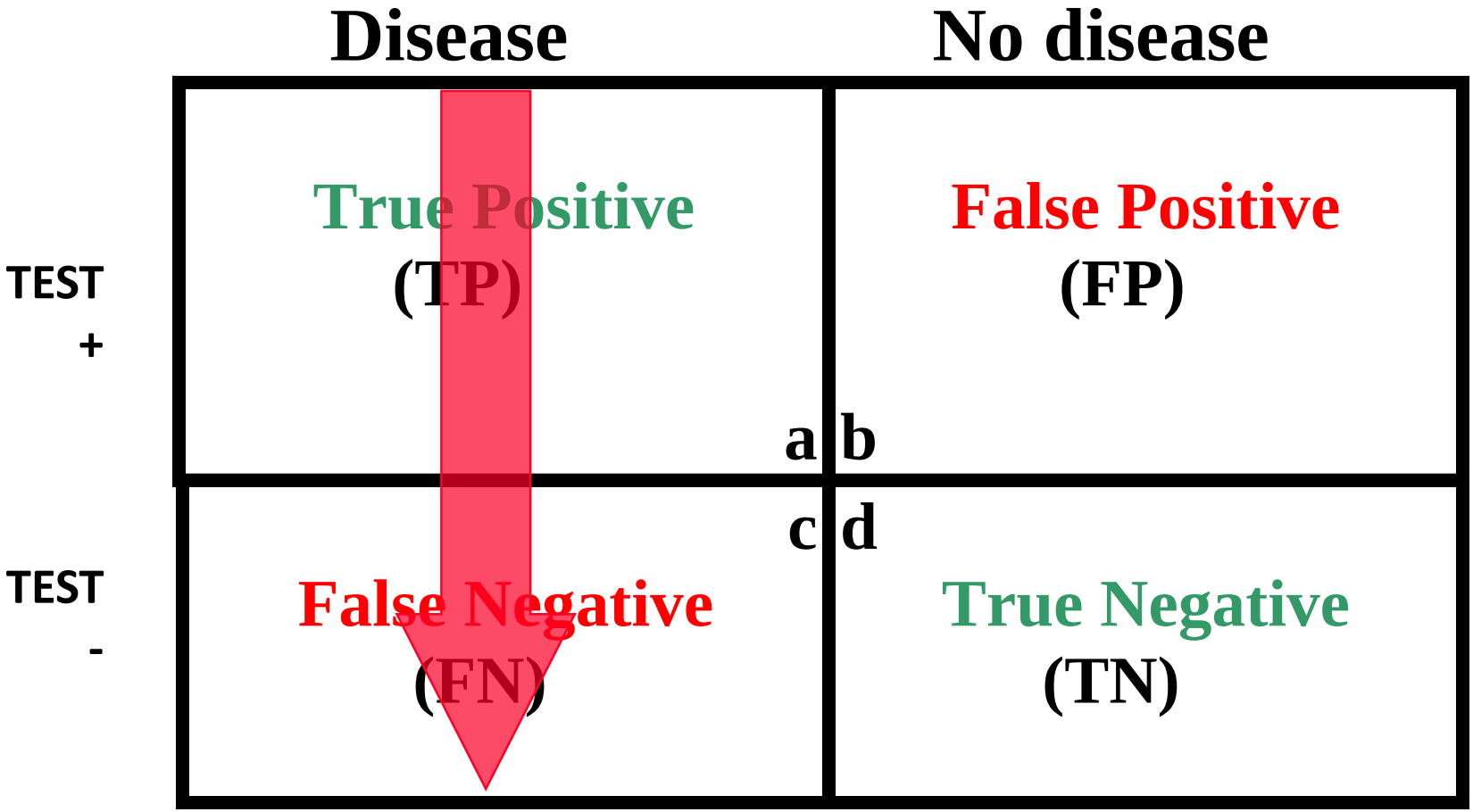
Post test 20%

? Appendicitis:

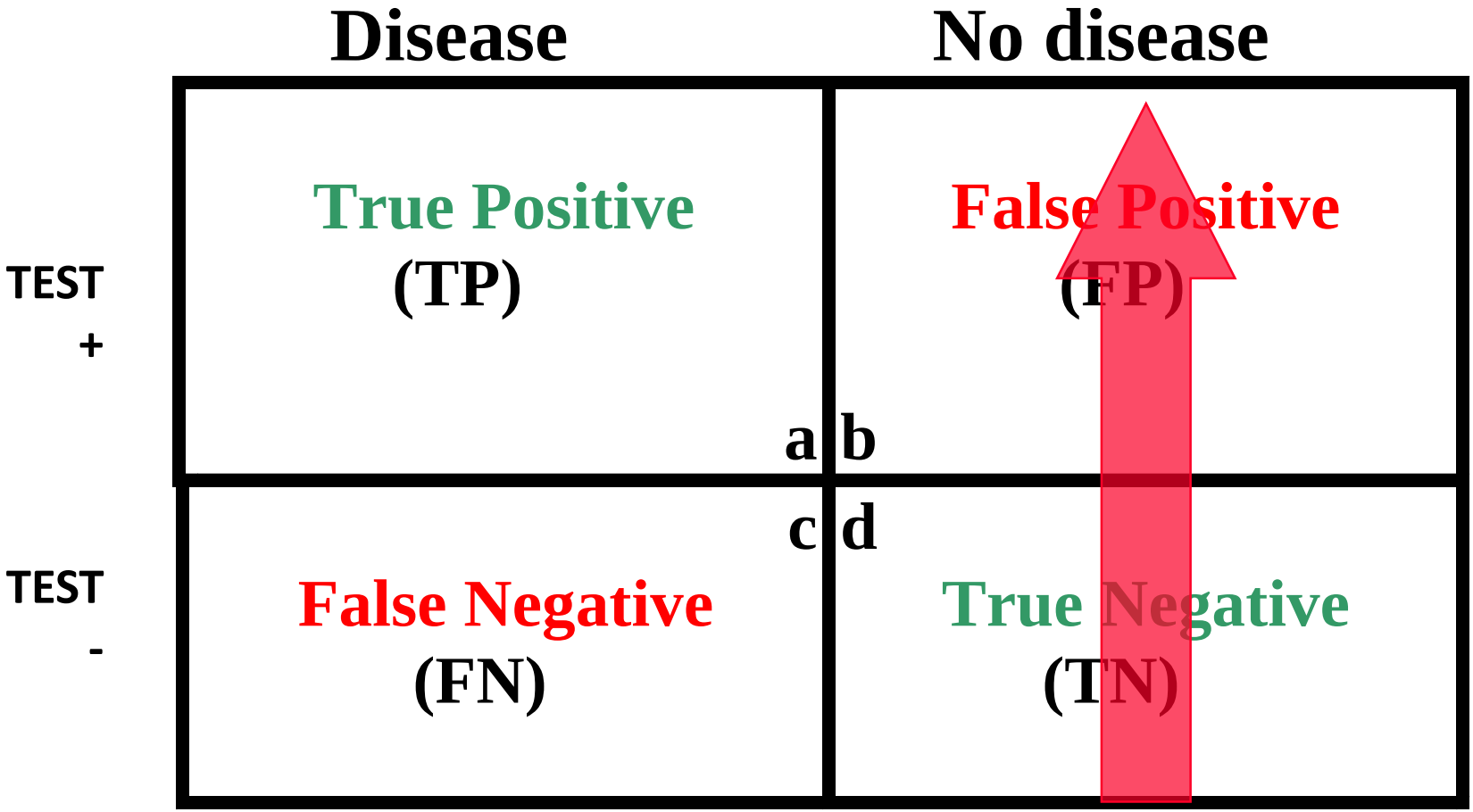
McBurney tenderness
LR+ = 3.4


Disease
No disease
TEST +
TEST -
True Positive (TP)
False Positive (FP)
a
b
c
d
False Negative (FN)
True Negative (TN)

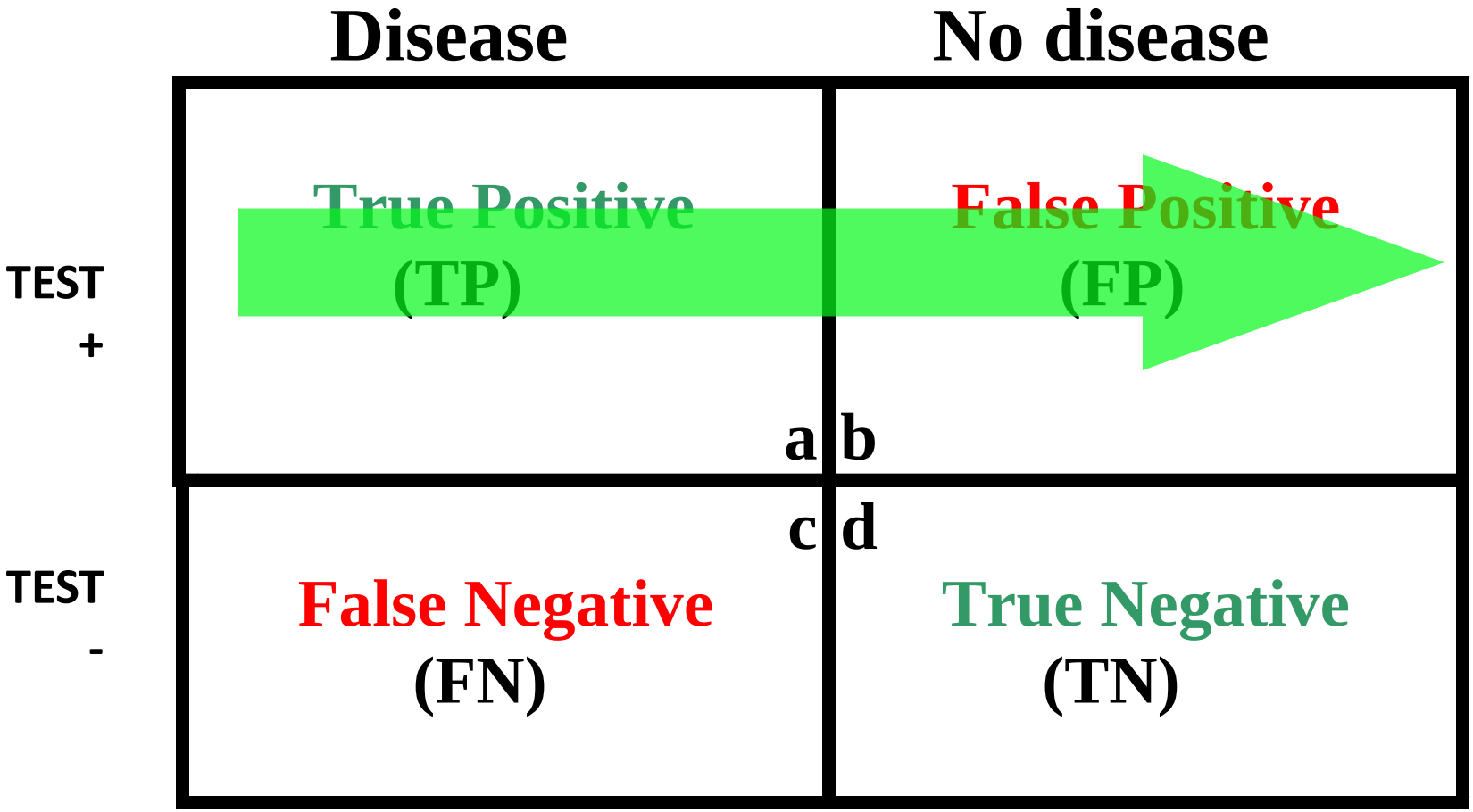
Sensitivity



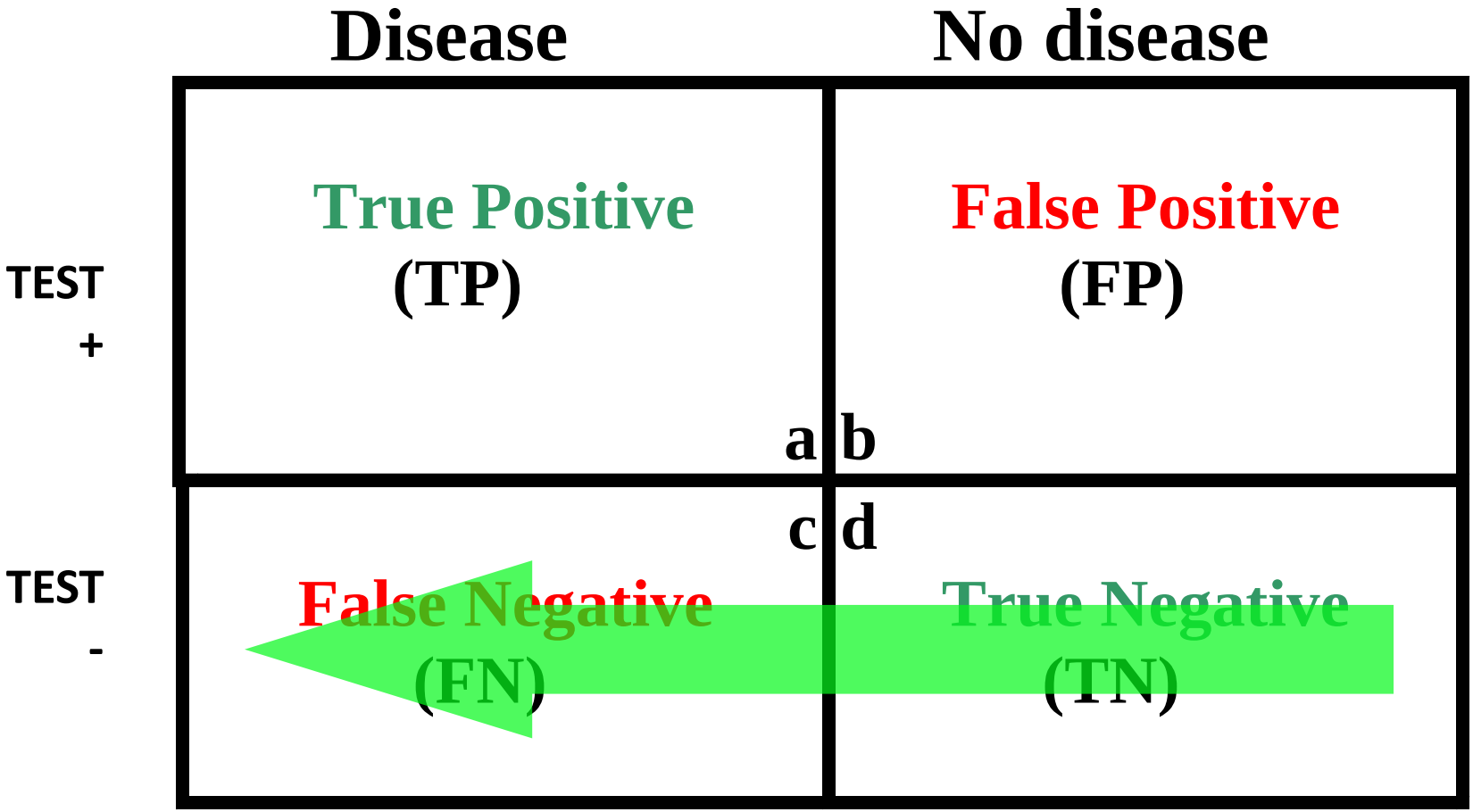
Specificity



Positive Predictive Value



Negative Predictive Value



Likelihood Ratios

- Similar to the concepts of “ruling in” and “ruling out” disease
- Pre Test Odds x LR = Post Test Odds
- The problem – we don’t think in terms of odds
- Clinical decision rules: Do the hard math for us, be we need to enter the appropriate data and interpret results

Levels of internal validity

1. Were there enough subjects in the study?
2. Was a control group used?
3. Were the subjects randomly assigned?
4. Was a pretest used?
5. Was the study started prior to the intervention or event?
6. Was the outcome measured in an objective and reliable way?

6x yes	= very high (A)
5x yes	= high (A)
4-3x yes	= limited (B)
2x yes	= low (C)
1-0x yes	= very low (D)

Appraisal

Critical appraisal questionnaires

www.cebma.org/ebp-tools

Appraisal of a cohort

1. Did the study address a clearly focused issue?
2. Was the cohort / panel recruited in an acceptable way? (selection bias)
3. Was the cohort/ panel representative of a defined population?
4. Was a control group used? Should one have been used?
5. Are objective and validated measurement methods used and were they similar in the different groups? (misclassification bias)
6. Was the follow up of cases/subjects long enough?
7. Could there be confounding?
8. Is the size of effect practically relevant?
9. Are the conclusions applicable?

Downloads:

•There are different worksheets available. Click on the title to access the Microsoft word (.doc) versions of the worksheets:

- Diagnosis worksheet
- Harm worksheet
- Prognosis worksheet
- Systematic review (of therapy) worksheet
- Therapy worksheet

https://ebm-tools.knowledgetranslation.net/worksheet

Mac iPad iPhone Watch TV Music Support

App Store Preview

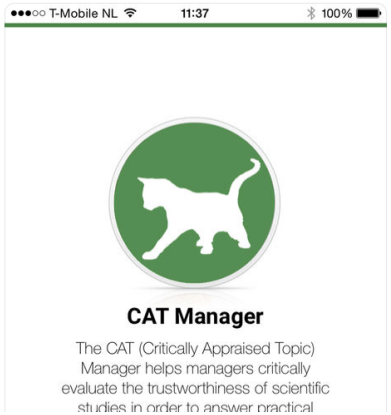
This app is only available on the App Store for iOS devices.

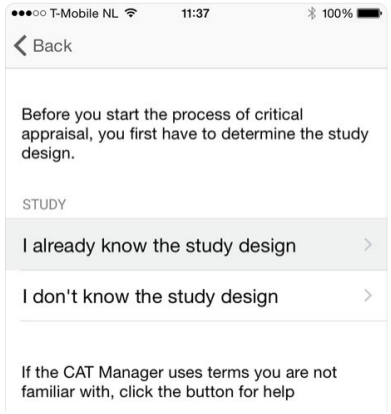
CAT Manager 4+

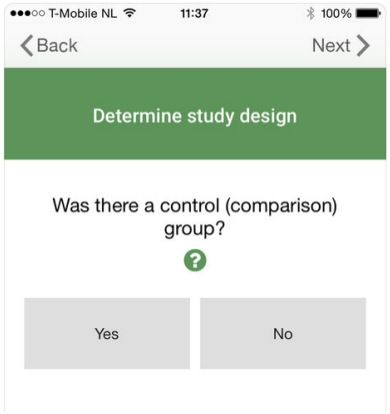
Center for Evidence-Based Management

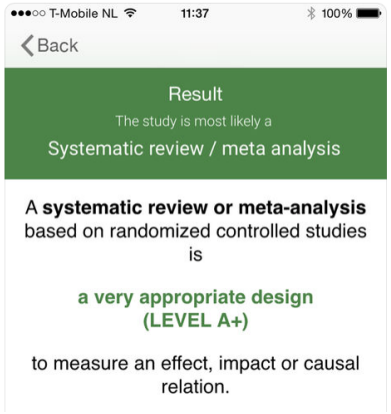
Free

iPhone Screenshots









- http://ktclearinghouse.ca/cebm/practise/ca/prognosis

Take-home messages:

- Different types of question require different study designs.
- Does the study address a clearly focused question?
- Did the study use valid methods to address this question?
- Are the valid results of this study important?
- Are these valid, important results applicable to my patient or population?

Refrences

- http://www.casp-uk.net/e-learning
- http://www.casp-uk.net/checklists
- http://clinicalevidence.bmj.com/x/set/static/ebm/toolbox/665061.html
- Critical Appraisal tools and Work sheets http://www.cebm.net/critical-appraisal/
- Jane M Young & Michael J Solomo. How to critically appraise an article. Nature Reviews Gastroenterology and Hepatology 6, 82-91 (February 2009) | doi:10.1038/ncpgasthep1331
- What is a critical appraisal? http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what_is_critical_appraisal.pdf
- The BMJ - How to Read a Paper http://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper
- Evidently Cochrane http://www.evidentlycochrane.net/
- CEBM http://www.cebm.net/

جامعة الملك سعود
King Saud University

Critical Appraisal (MA/SR/Guidelines)

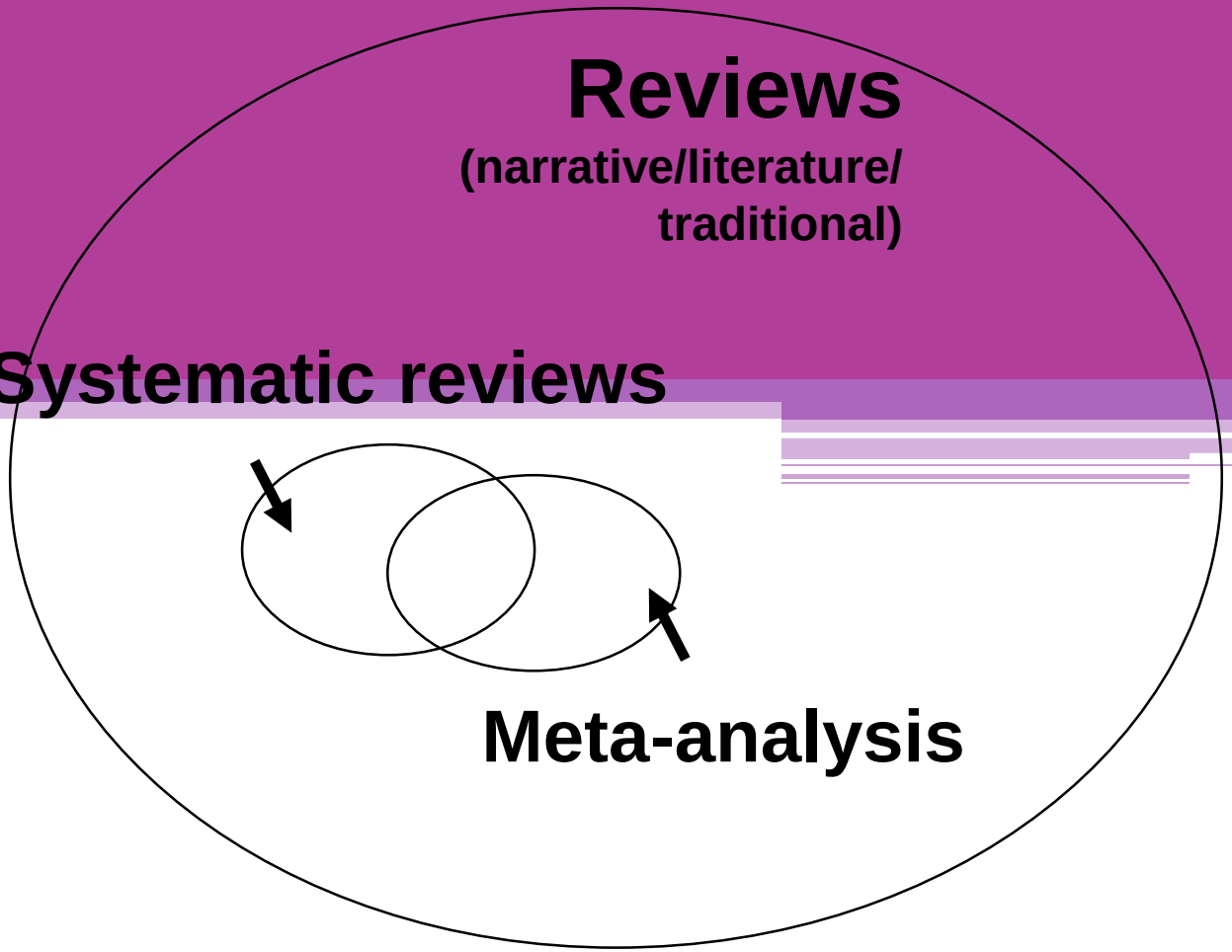
Dr. Nada A. AlYousefi

Associate Professor, Postgraduate Trainer and Consultant of Family Medicine
International Board Certified Lactation Consultant
Department of Family and Community Medicine
College of Medicine, King Saud University (KSU)

nalyousefi@ksu.edu.sa

January 2019

Types of reviews



Narrative reviews

- Usually written by experts in the field
- Use informal and subjective methods to collect and interpret information
- Usually narrative summaries of the evidence

Read: Klassen et al. Guides for Reading and Interpreting Systematic Reviews. Arch Pediatr Adolesc Med 1998;152:700-704.

What is a systematic review?

A review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyse data from the studies that are included in the review*

*Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD Report Number 4 (2nd Edition). NHS Centre for Reviews and Dissemination, University of York. March 2001.

Key elements of a systematic review

Structured, systematic process involving several steps :

1. Formulate the question
2. Plan the review
3. Comprehensive search
4. Unbiased selection and abstraction process
5. Critical appraisal of data
6. Synthesis of data (may include meta-analysis)
7. Interpretation of results

All steps described explicitly in the review

Limitations of systematic reviews specific to health promotion

- Results may still be inconclusive
- There may be no trials/evidence
- The trials may be of poor quality
- The intervention may be too complex to be tested by a trial
- Practice does not change just because you have the evidence of effect/effectiveness

EBM and Systematic Review

- EBM

Steps

1. Question (PICO)?
2. Find the best evidence?
3. Appraise?
4. Synthesised?
5. Apply?

Time: 120 seconds

1 - 20 articles

This patient survives!

- Systematic Review

Steps

1. Question (PICO)
2. Find the best evidence x 2+
3. Appraise x 2+
4. Synthesize
5. ---

Time: 6 months+, team

≤ 2,000 articles

This patient is dead

Find a systematic review (and appraise it quickly)!

The Cochrane Collaboration

International non-profit organisation that prepares, maintains, and disseminates systematic up-to-date reviews of health care interventions

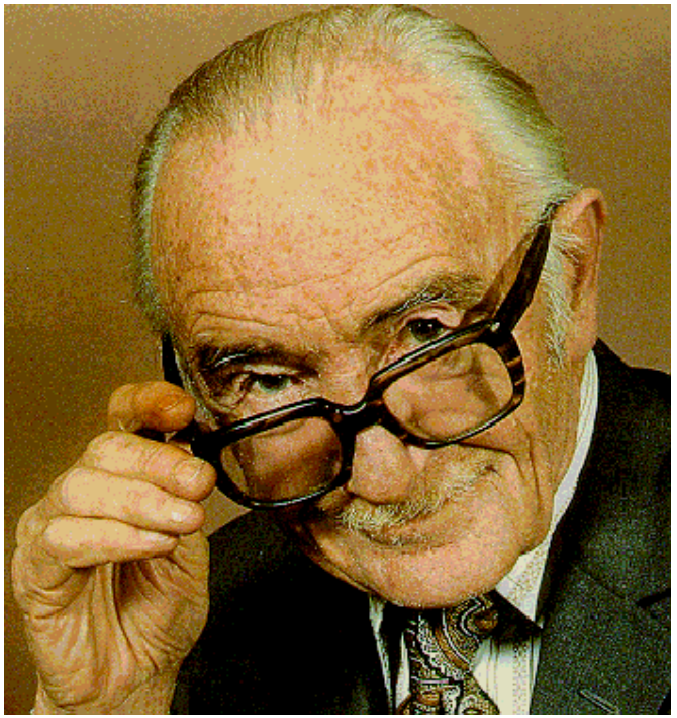
1992

Cochrane Collaboration

Named in honour of Archie Cochrane, a British researcher

In 1979:

"It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials"



The Cochrane Library

- **Cochrane Systematic reviews :** Cochrane reviews and protocols
- **Database of Reviews of Effects:** Other systematic reviews appraised by the Centre for Reviews and Dissemination.
- **Cochrane Central Register of Controlled Trials:** Bibliography of controlled trials (some not indexed in MEDLINE).
- **Health Technology Assessment Database:** HTA reports
- **NHS Economic evaluation database:** Economic evaluations of health care interventions.

Appraising a systematic review

Knee Surg Sports Traumatol Arthrosc (2010) 18:304–311
DOI 10.1007/s00167-009-0965-z

KNEE

Early versus delayed surgery for anterior cruciate ligament reconstruction: a systematic review and meta-analysis

Toby O. Smith · Leigh Davies · Caroline B. Hing

Received: 1 July 2009 / Accepted: 5 October 2009 / Published online: 17 October 2009

Abstract There is no consensus in the literature regarding the optimal timing of surgical reconstruction of the ruptured anterior cruciate ligament (ACL). Previous authors have suggested that early reconstruction may facilitate an early return to work or sport but may increase the incidence of post-operative complications such as arthrofibrosis. This study systematically reviewed the literature to determine whether ACL reconstruction should be performed acutely following rupture. Medline, CINAHL, AMED, EMBASE databases and grey literature were reviewed with a meta-analysis of pooled mean differences where appropriate. Six papers including 370 ACL reconstructions were included. Early ACL reconstructions were considered as those undertaken within a mean of 3 weeks post-injury; delayed ACL reconstructions were those undertaken a minimum of 6 weeks post-injury. We found there was no difference in clinical outcome between patients who underwent early compared to delayed ACL reconstruction. However, this conclusion is based on the current literature which has substantial methodological limitations.

Keywords Anterior cruciate ligament · Reconstruction · Timing of surgery · Meta-analysis

T. O. Smith (✉)
Institute of Orthopaedics, Norfolk and Norwich University Hospital, Colney Lane, Norwich, Norfolk NR2 7UY, UK
e-mail: toby.smith@nnuh.nhs.uk

T. O. Smith
University of East Anglia, Norwich, UK

L. Davies
Physiotherapy Department, Norfolk and Norwich University Hospital, Norwich, UK

C. B. Hing
Watford General Hospital, Watford, UK

Introduction

The anterior cruciate ligament (ACL) is the most frequently injured ligament of the knee with an incidence of 8 per 100,000 cases per year [6, 28]. Surgery is the typical treatment for younger athletes or those with physically demanding occupational or sporting pursuits since it restores stability and limits the potential for progressive degeneration and long-term instability of the knee [2, 4, 19].

Surgical techniques of ACL reconstruction have evolved over the past three decades with debate regarding timing of reconstruction [37]. In a national survey by Francis et al. [12], of 101 consultant orthopaedic surgeons in the UK, 81% reported that they considered the ideal time span from injury to operation to be between 1 and 6 months, although it was acknowledged that only 35% of ACL reconstructions are performed within this time-frame in National Health Service hospitals.

Proponents of early surgical intervention during the initial weeks post-injury have suggested that restoring tibiofemoral stability may minimise the risk of further meniscal and chondral injury which may be associated with degenerative joint changes [3, 9, 35]. Early surgery may also facilitate return to sporting and occupational pursuits with considerable economic consequences. Delayed ACL reconstruction may be associated with an increase in muscle atrophy and reduced strength which may delay early rehabilitation [10, 29]. Conversely, delaying surgical intervention allows optimisation of pre-operative knee range of motion and recovery of surrounding soft tissues from the initial injury potentially reducing the incidence of

?

Tools for critical appraisal

- CASP: Critical Appraisal Skills Programme Checklists
- Critically Appraised Topics: generic systematic reviews (ACP Journal club)
- SIGN: Scottish Intercollegiate Guidelines Network
- GATE Frame

Critically Appraising the E | www.cebm.net/wp-conte

www.cebm.net/category/ebm-resources/tools/critically-appraising-the-evidence/

Apps | WebLearn : Welcom... | Posts ‹ David Nunan... | David Nunan | ab

CEBM

CENTRE FOR EVIDENC

HOME | EDUCATION & TRAINING | EVENTS

Critically Appraising th

Evaluation a report of a study to determine whether it is valid, importa

Home > EBM Resources > Tools > Critically Appraising the Evidence

Critical Appraisal tools

Critical appraisal worksheets to help you appraise the reliability,

Critical Appraisal Worksheets

English

- Systematic Review Critical Appraisal Sheet
- Diagnosis Critical Appraisal Sheet
- Prognosis Critical Appraisal Sheet
- Therapy / RCT Critical Appraisal Sheet

German – Translated by Johannes Pohl and Martin Sadilek

- Systematic Review German Translation (PDF)
- Diagnosis German Translation (PDF)
- Prognosis German Translation (PDF)
- Therapy / RCT German Translation (PDF)

Spanish – Translated by Ana Cristina Castro

- Systematic Review (PDF)
- Diagnosis (PDF)
- Prognosis Spanish Translation (PDF)
- Therapy / RCT Spanish Translation (PDF)

Lithuanian – Translated by Tumas Beinortas

- Systematic review appraisal Lithuanian (PDF)
- Diagnostic accuracy appraisal Lithuanian (PDF)
- Prognostic study appraisal Lithuanian (PDF)
- RCT appraisal sheets Lithuanian (PDF)

Appraisal of a systematic review

- 10 questions
  1. Clearly-focused question
  2. The right type of study included
  3. Identifying all relevant studies
  4. Assessment of quality of studies
  5. Reasonable to combine studies
  6. What were the results
  7. Preciseness of results
  8. Application of results to local population
  9. Consideration of all outcomes
  10. Policy or practice change as a result of evidence

CASP

Appraising a systematic review

Knee Surg Sports Traumatol Arthrosc (2010) 18:304–311
DOI 10.1007/s00167-009-0965-z

KNEE

Early versus delayed surgery for anterior cruciate ligament reconstruction: a systematic review and meta-analysis

Toby O. Smith · Leigh Davies · Caroline B. Hing

Received: 1 July 2009 / Accepted: 5 October 2009 / Published online: 17 October 2009

Abstract There is no consensus in the literature regarding the optimal timing of surgical reconstruction of the ruptured anterior cruciate ligament (ACL). Previous authors have suggested that early reconstruction may facilitate an early return to work or sport but may increase the incidence of post-operative complications such as arthrofibrosis. This study systematically reviewed the literature to determine whether ACL reconstruction should be performed acutely following rupture. Medline, CINAHL, AMED, EMBASE databases and grey literature were reviewed with a meta-analysis of pooled mean differences where appropriate. Six papers including 370 ACL reconstructions were included. Early ACL reconstructions were considered as those undertaken within a mean of 3 weeks post-injury; delayed ACL reconstructions were those undertaken a minimum of 6 weeks post-injury. We found there was no difference in clinical outcome between patients who underwent early compared to delayed ACL reconstruction. However, this conclusion is based on the current literature which has substantial methodological limitations.

Keywords Anterior cruciate ligament · Reconstruction · Timing of surgery · Meta-analysis

Introduction

The anterior cruciate ligament (ACL) is the most frequently injured ligament of the knee with an incidence of 8 per 100,000 cases per year [6, 28]. Surgery is the typical treatment for younger athletes or those with physically demanding occupational or sporting pursuits since it restores stability and limits the potential for progressive degeneration and long-term instability of the knee [2, 4, 19].

Surgical techniques of ACL reconstruction have evolved over the past three decades with debate regarding timing of reconstruction [37]. In a national survey by Francis et al. [12], of 101 consultant orthopaedic surgeons in the UK, 81% reported that they considered the ideal time span from injury to operation to be between 1 and 6 months, although it was acknowledged that only 35% of ACL reconstructions are performed within this time-frame in National Health Service hospitals.

Proponents of early surgical intervention during the initial weeks post-injury have suggested that restoring tibiofemoral stability may minimise the risk of further meniscal and chondral injury which may be associated with degenerative joint changes [3, 9, 35]. Early surgery may also facilitate return to sporting and occupational pursuits with considerable economic consequences. Delayed ACL reconstruction may be associated with an increase in muscle atrophy and reduced strength which may delay early rehabilitation [10, 29]. Conversely, delaying surgical intervention allows optimisation of pre-operative knee range of motion and recovery of surrounding soft tissues from the initial injury potentially reducing the incidence of

T. O. Smith (✉)
Institute of Orthopaedics, Norfolk and Norwich University Hospital, Colney Lane, Norwich, Norfolk NR2 7UY, UK
e-mail: toby.smith@nnuh.nhs.uk

T. O. Smith
University of East Anglia, Norwich, UK

L. Davies
Physiotherapy Department, Norfolk and Norwich University Hospital, Norwich, UK

C. B. Hing
Watford General Hospital, Watford, UK

3 minutes

Step 1 – Are the results of the review valid?

- **Q**uestion – what is the PICO (etc.)
- **F**inding – comprehensive?
- **A**ppraise – did they select good ones?
- **S**ynthesise – numerically/appropriate?

1. What question (PICO) did the systematic review address?

- Is question clearly stated early on?
- Treatment/exposure described?
- Comparator/control described?
- Outcome(s) described?

Title, abstract, introduction

Knee Surg Sports Traumatol Arthrosc (2010) 18:304–311

post-operative arthrofibrosis and wound complications [17, 31, 37, 38].

There is no consensus in the current literature regarding the optimal time of surgical intervention [29]. The purpose of this study was to assess the effects of duration from injury to surgical intervention for patients undergoing ACL reconstruction by comparing the clinical and radiological outcomes of early to delayed ACL reconstruction following initial injury.

P

I

C

O's

2. Is it unlikely that important, relevant studies were missed?

Look for

- Which bibliographic databases were used? More than 1?
- Search terms used (text and MeSH)?
- Search for unpublished as well as published studies?
- Search for non-English studies?

Methods

Is finding all published studies enough?

- Negative studies less likely to be published than 'Positive' ones
- How does this happen?
- Positive studies SUBMITTED 2.5x more often than negative (Dickersin, JAMA, 1992)

Patients and methods

Data sources and searches

A database search was performed via Ovid of Medline (1950 to June 2009), CINAHL (1982 to June 2009), AMED (1985 to June 2009) and EMBASE (1974 to June 2009) using MeSH terms to identify all English-language randomised and non-randomised clinical trials specifically comparing outcomes of early versus delayed ACL reconstructions. The key word terms and Boolean operators used were "anterior cruciate ligament reconstruction" AND "surgery" AND "timing" OR "delay." We also searched for unpublished literature using the search term "anterior cruciate ligament" from the databases SIGLE (System for Information on Grey Literature in Europe), the National Technical Information Service, the National Research Register (UK) and Current Controlled Trials databases. We attempted to contact the corresponding authors of each included paper to highlight any omitted citations. Trials

3. Were the criteria used to select articles for inclusion appropriate?

Look for

- Inclusion/exclusion criteria a priori?
- Are eligibility criteria related to PICO?
- Types of studies?

Methods

or arthroscopic, the type of graft, gender or post-operative rehabilitation. The reference lists of review papers were scrutinised for relevant publications not identified by the initial search strategy. Single case reports, comments, letters, editorials, protocols, guidelines and review papers were excluded. We also excluded studies evaluating cases under the age of 16; studies of revision ACL reconstruction; studies presenting result of ACL repair rather than reconstruction; and papers which did not specifically detail the range of time between injury and surgery for their acute and delayed groups. Two investigators (TS, LD) independently selected articles meeting the

Patients and methods

Data sources and searches

A database search was performed via Ovid of Medline (1950 to June 2009), CINAHL (1982 to June 2009), AMED (1985 to June 2009) and EMBASE (1974 to June 2009) using MeSH terms to identify all English-language randomised and non-randomised clinical trials specifically comparing outcomes of early versus delayed ACL recon-

4. Were the included studies sufficiently valid for the type of question?

Look for

- Criteria for quality assessment defined?
- Appropriate for the question?
- Were the assessment results provided?

Methods, Results

Criteria for quality assessment defined?

Data extraction and quality assessment

Two investigators (TS, LD), blinded to the source, publication date, authors and affiliations for each paper, used a standardised extraction form. All papers were then evaluated against the eleven-item PEDro scoring system by TS and LD independently. The PEDro appraisal tool has demonstrated reliability and validity in the assessment of

Appropriate for the question?

Were assessment results provided?

Table 3 PEDro critical appraisal results

	Bottoni et al. [4]	Marcacci et al. [26]	Meighan et al. [28]	Petersen and Laprell [34]	Sgaglione et al. [35]	Wasilewski et al. [42]
Eligibility criteria	1	0	1	0	1	0
Random allocation	1	0	1	0	0	0
Concealed allocation	1	0	0	0	0	0
Baseline comparability	1	0	0	0	0	1
Blind subject	0	0	0	0	0	0
Blind clinician	0	0	0	0	0	0
Blind assessor	0	0	1	0	0	0
Adequate follow-up	1	1	1	0	1	1
Intention-to treat analysis	0	0	1	0	0	0
Between-group analysis	1	1	1	1	1	1
Point estimates and variability	1	0	0	1	1	0
Total score	7	2	6	2	4	3

1 one point, 0 no point

5. Were the results similar from study to study?

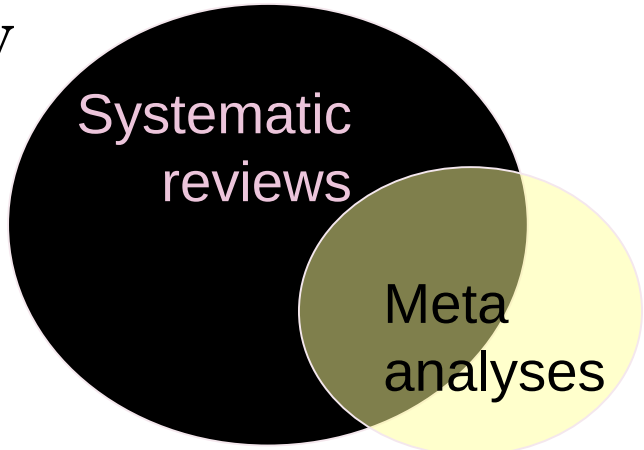
Consider whether

- The results of all the included studies are clearly displayed
- The results are combined (meta-analysis)
  - *Are studies sufficiently similar*
- The reasons for any variations in results are discussed

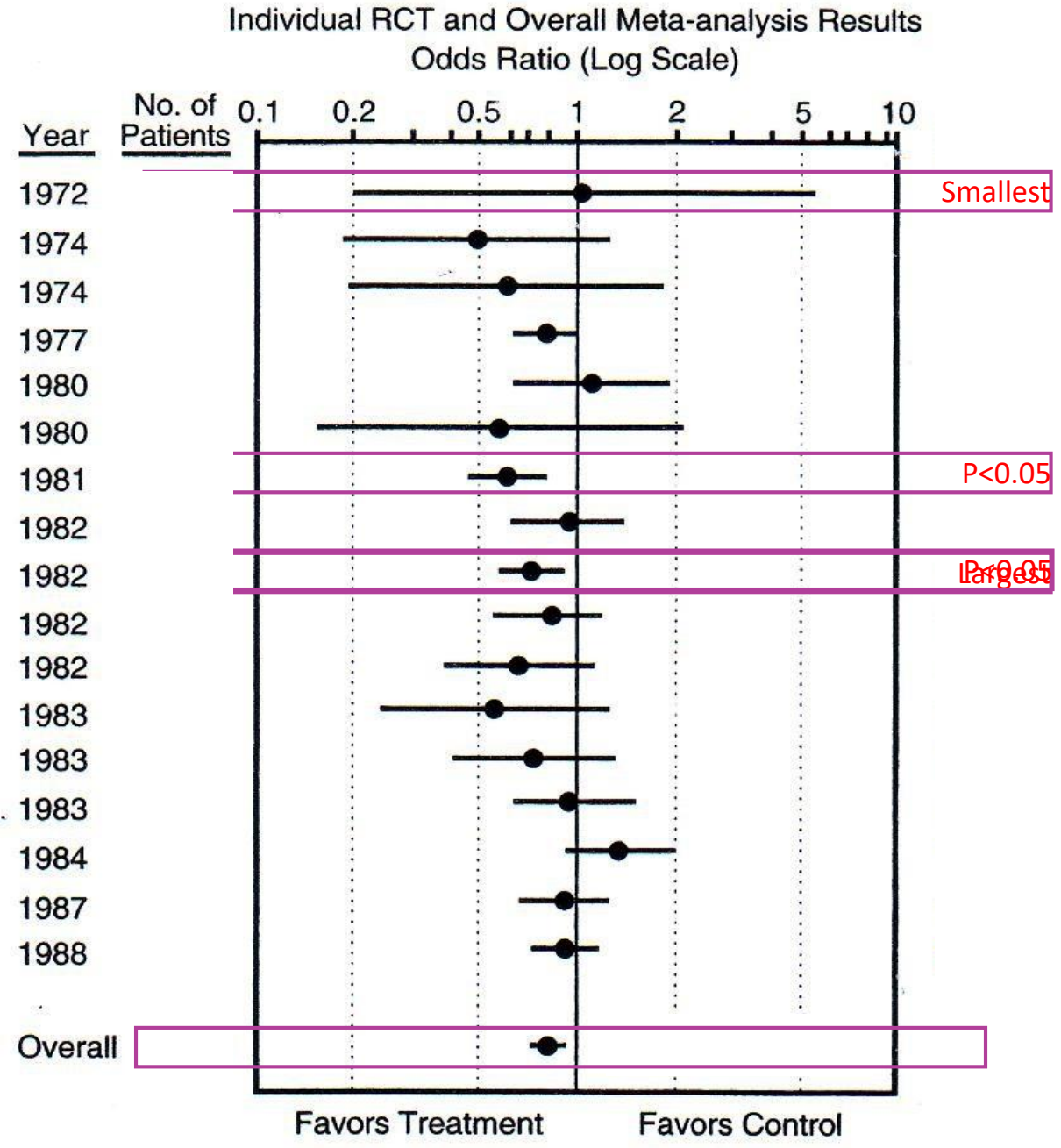
Meta-analysis

= calculated "best guess" of the true effect size

- The statistical combination of the results gives a pooled, weighted average of the primary results
- It weights the effect size (result) of each study in relation to sample size of the study
- Optional part of SR



FOREST PLOTS



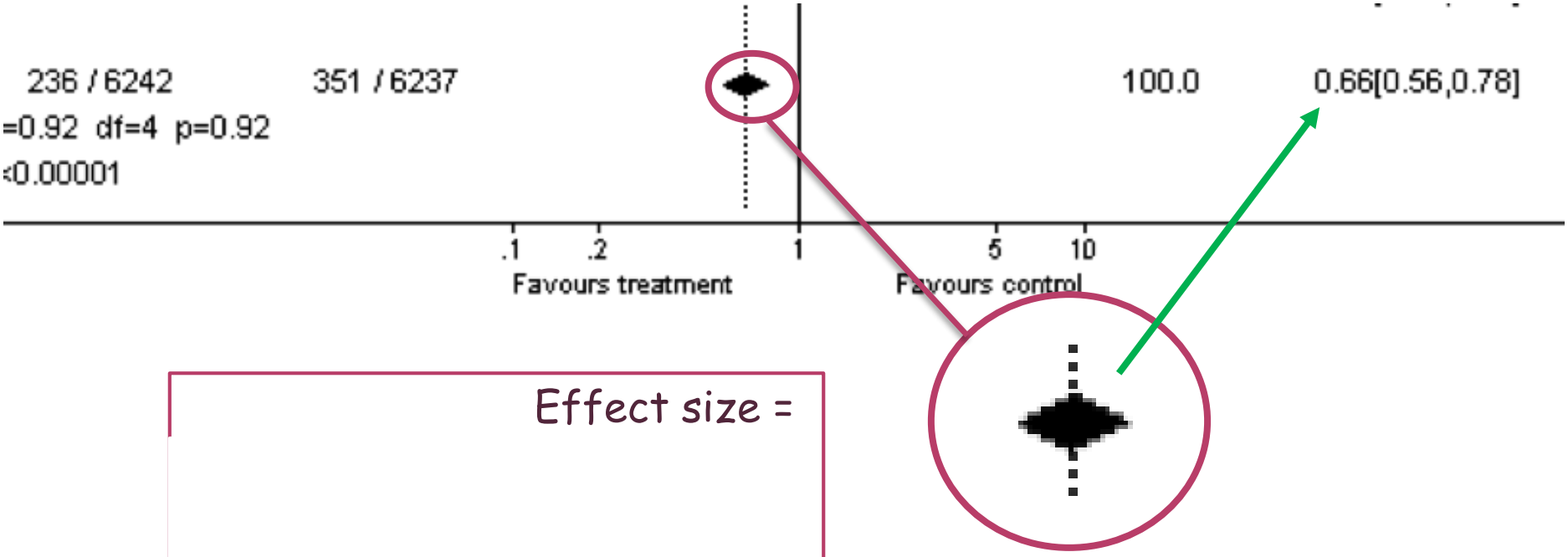
Which is the smallest study? .A

Which is the largest study? .B

How many are statistically significant? .C

Is treatment better than control?

How much better?


236 / 6242
351 / 6237
100.0
0.66[0.56,0.78]
=0.92 df=4 p=0.92
<0.00001
.1
.2
1
5
10
Favours treatment
Favours control

Effect size =

Heterogeneity

"The quality or state of being diverse in character or content"

Heterogeneity (diversity)

- **Clinical heterogeneity**

Variability in the participants, interventions and/or outcomes studied

- **Methodological heterogeneity**

Variability in study deign and risk of bias

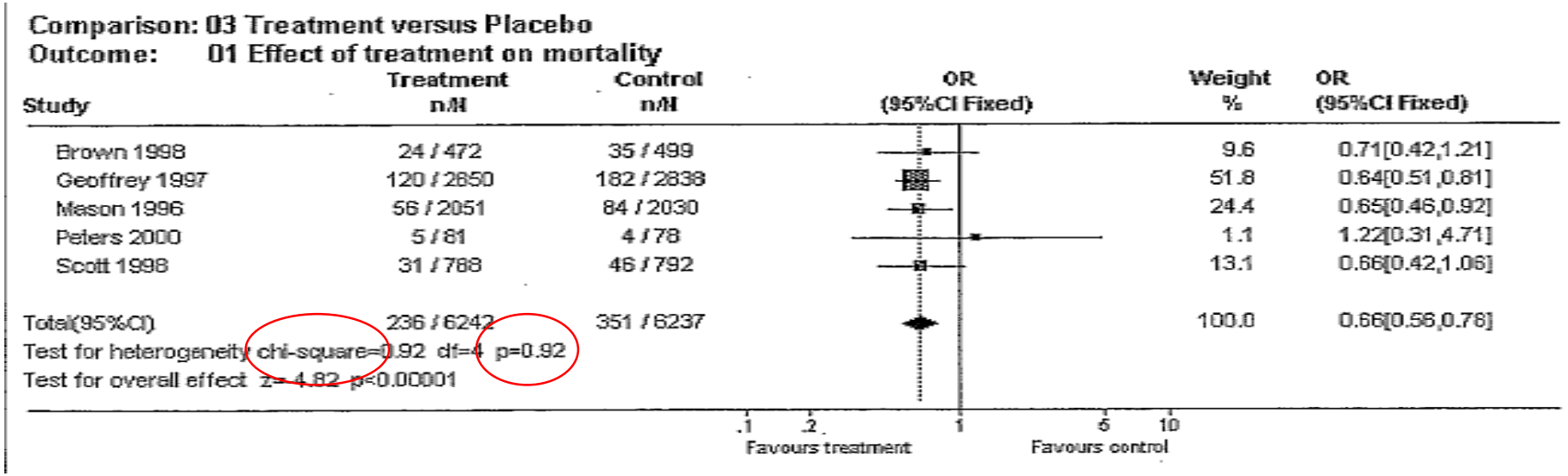
- **Statistical heterogeneity**

The observed intervention effects being more different from each other than we would expect due to random error (chance) alone

High heterogeneity

=

appropriate to pool data?



Are the results <u>similar</u> across studies? 3 tests:

1. 'Eyeball' test – do they look they same?

2. Formal tests
   a) Proportion of variation not due to chance (I^2)
      - Variation between tests, now preferred estimate
        - 0% to 40%: might not be important;
        - 30% to 60%: may represent moderate heterogeneity;
        - 50% to 90%: may represent substantial heterogeneity;
        - 75% to 100%: considerable heterogeneity

   b) Test of 'Null hypothesis' of no variation (p-value)
      - Cochrane Chi-square: $p<0.10$ = heterogeneity

Are these trials different?

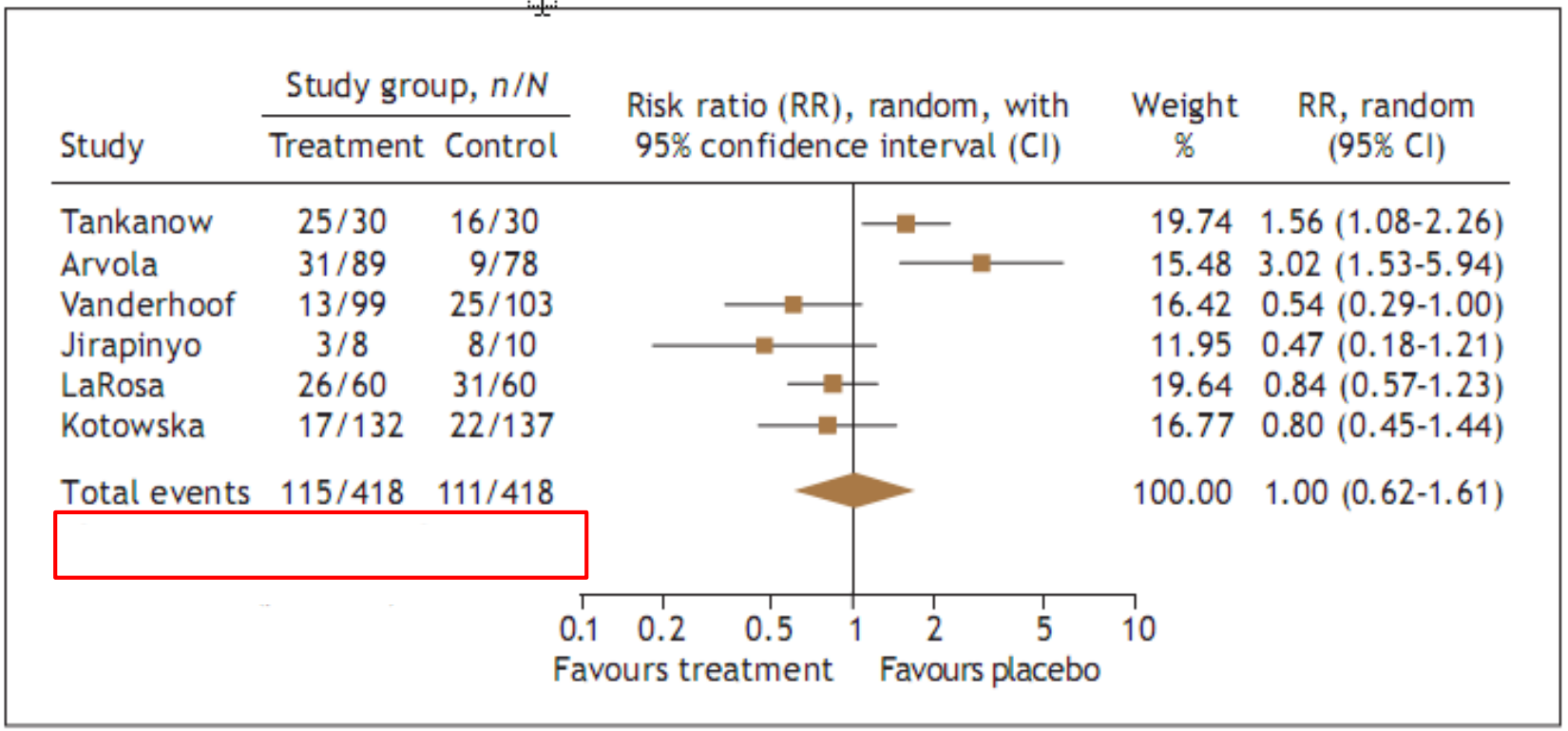


Fig 3: Incidence of antibiotic-associated diarrhea — intention-to-treat analysis. The analysis showed a nonsignificant difference between probiotics and placebo (z score) and statisti-

5. Were the results similar from study to study?

Table 2 Results of meta-analysis

Outcome	Papers	Relative risk (95% CI)	Overall effect (*P* value)	Heterogeneity	
				χ^2	I^2
Lysholm Score	[4, 34, 35]	0.07 (−9.93, 10.08)*	0.99	0.02	81
Lysholm Score (Good/excellent)	[26]				
Tegner Score	[4, 34, 35]	−0.07 (−0.42, 0.29)*	0.71	0.60	0
KT-1000 Arthrometer	[4, 34, 35]	0.05 (−0.52, 0.63)*	0.85	0.19	42
Tibiofemoral Displacement > 3 mm	[25, 35]	0.59 (0.25, 1.43)	0.24	0.19	43
Positive Lachman	[26, 34, 35]	0.64 (0.27, 1.51)	0.31	0.02	73
Positive pivot shift	[26, 34, 35]	0.69 (0.43, 1.11)	0.13	0.52	0
Extension deficit	[4, 35]	−0.90 (−2.39, 0.59)*	0.24	N/E	N/E
Flexion deficit	[4, 35]	−0.50 (−2.55, 1.55)*	0.63	N/E	N/E
Extension deficit > 10°	[4, 26, 34]	0.96 (0.21, 4.37)	0.96	0.21	36
Incidence of arthrofibrosis	[28, 34, 35, 42]	1.83 (0.81, 4.14)	0.15	0.76	0
Incidence of meniscal injury	[4, 26, 28, 34, 42]	0.92 (0.71, 1.19)	0.53	<0.01	74
Incidence of chondral injury	[4, 26, 34, 42]	0.77 (0.44, 1.37)	0.38	0.26	25
Frequency of revision surgery	[26, 28, 34, 35, 42]	0.81 (0.42, 1.58)	0.54	0.30	17
Incidence of patellofemoral pain	[35, 42]	2.05 (0.86, 4.89)	0.11	0.58	0
Incidence of thromboembolic complication	[28, 35]	1.79 (0.21, 27.29)	0.68	0.21	37

* Mean difference (95% confidence intervals), ° degrees, *CI* confidence intervals, *mm* millimetres, N/E not estimated

Step 2 – What were the results?

Consider

- What these are (numerically if appropriate)
- How were the results presented/expressed (risk ratio, odds ratio, etc.)
- If you are clear about the review's 'bottom line' results

What are we interested in?

Table 2 Results of meta-analysis

Outcome	Papers	Relative risk (95% CI)	Overall effect (*P* value)	Heterogeneity	
				χ^2	I^2
Lysholm Score	[4, 34, 35]	0.07 (−9.93, 10.08)*	0.99	0.02	81
Lysholm Score (Good/excellent)	[26]				
Tegner Score	[4, 34, 35]	−0.07 (−0.42, 0.29)*	0.71	0.60	0
KT-1000 Arthrometer	[4, 34, 35]	0.05 (−0.52, 0.63)*	0.85	0.19	42
Tibiofemoral Displacement > 3 mm	[25, 35]	0.59 (0.25, 1.43)	0.24	0.19	43
Positive Lachman	[26, 34, 35]	0.64 (0.27, 1.51)	0.31	0.02	73
Positive pivot shift	[26, 34, 35]	0.69 (0.43, 1.11)	0.13	0.52	0
Extension deficit	[4, 35]	−0.90 (−2.39, 0.59)*	0.24	N/E	N/E
Flexion deficit	[4, 35]	−0.50 (−2.55, 1.55)*	0.63	N/E	N/E
Extension deficit > 10°	[4, 26, 34]	0.96 (0.21, 4.37)	0.96	0.21	36
Incidence of arthrofibrosis	[28, 34, 35, 42]	1.83 (0.81, 4.14)	0.15	0.76	0
Incidence of meniscal injury	[4, 26, 28, 34, 42]	0.92 (0.71, 1.19)	0.53	<0.01	74
Incidence of chondral injury	[4, 26, 34, 42]	0.77 (0.44, 1.37)	0.38	0.26	25
Frequency of revision surgery	[26, 28, 34, 35, 42]	0.81 (0.42, 1.58)	0.54	0.30	17
Incidence of patellofemoral pain	[35, 42]	2.05 (0.86, 4.89)	0.11	0.58	0
Incidence of thromboembolic complication	[28, 35]	1.79 (0.21, 27.29)	0.68	0.21	37

* Mean difference (95% confidence intervals), ° degrees, *CI* confidence intervals, *mm* millimetres, N/E not estimated

Our clinical question

Amongst adults with acute ACL injuries, does

early reconstructive surgery compared with

delayed reconstructive surgery lead to

favourable return to former activity and/or risk

of recurrent knee injury?

Population

Intervention

Control

Outcome 1

Outcome 2

Return to former activity (page 306):

There was no statistically significant difference between the early and delayed ACL reconstruction groups for the Lysholm score or Tegner score (Table 2). There was no significant difference between the groups for International Knee Documentation Committee rating score [not significant (n.s.)] [26], IKDC perceived stability rating (n.s.) [26], or the Hospital for Special Surgery score system (n.s.) [35]. There was no reported significant difference in patient satisfaction ($P = 0.19$) [35]. The frequency that patients returned to the same level of sporting participation was assessed in Marcacci et al.'s [26] paper. This reported that there was no statistically significant difference in return rates between the two groups (n.s.) [26].

Risk of recurrent knee injury

Table 2 Results of meta-analysis

Outcome	Papers	Relative risk (95% CI)	Overall effect (*P* value)	Heterogeneity	
				χ^2	I^2
Lysholm Score	[4, 34, 35]	0.07 (−9.93, 10.08)*	0.99	0.02	81
Lysholm Score (Good/excellent)	[26]				
Tegner Score	[4, 34, 35]	−0.07 (−0.42, 0.29)*	0.71	0.60	0
KT-1000 Arthrometer	[4, 34, 35]	0.05 (−0.52, 0.63)*	0.85	0.19	42
Tibiofemoral Displacement > 3 mm	[25, 35]	0.59 (0.25, 1.43)	0.24	0.19	43
Positive Lachman	[26, 34, 35]	0.64 (0.27, 1.51)	0.31	0.02	73
Positive pivot shift	[26, 34, 35]	0.69 (0.43, 1.11)	0.13	0.52	0
Extension deficit	[4, 35]	−0.90 (−2.39, 0.59)*	0.24	N/E	N/E
Flexion deficit	[4, 35]	−0.50 (−2.55, 1.55)*	0.63	N/E	N/E
Extension deficit > 10°	[4, 26, 34]	0.96 (0.21, 4.37)	0.96	0.21	36
Incidence of arthrofibrosis	[28, 34, 35, 42]	1.83 (0.81, 4.14)	0.15	0.76	0
Incidence of meniscal injury	[4, 26, 28, 34, 42]	0.92 (0.71, 1.19)	0.53	<0.01	74
Incidence of chondral injury	[4, 26, 34, 42]	0.77 (0.44, 1.37)	0.38	0.26	25
Frequency of revision surgery	[26, 28, 34, 35, 42]	0.81 (0.42, 1.58)	0.54	0.30	17
Incidence of patellofemoral pain	[35, 42]	2.05 (0.86, 4.89)	0.11	0.58	0
Incidence of thromboembolic complication	[28, 35]	1.79 (0.21, 27.29)	0.68	0.21	37

* Mean difference (95% confidence intervals), ° degrees, *CI* confidence intervals, *mm* millimetres, N/E not estimated

What's the 'bottom line' of the review?

Conclusions

The findings of this study suggested that there was no statistically significant difference in outcomes between those patients who underwent earlier compared to delayed ACL reconstruction. The present evidence-base presented with substantial methodological limitations. A sufficiently powerful, well-design randomised controlled trial is required to determine whether of duration from injury to surgical intervention is an important prognostic indicator for patients who undergo an ACL reconstruction.

Can I apply these results to my case?

- Is my patient so different to those in the study that the results cannot apply?

early were compared to 209 delayed procedures. The mean age was 25.6 years in the early group [Standard deviation (SD) = 2.3] compared to 26.2 years (SD = 1.1) in the delayed group (Table 1).

‘Clinical pearls’

- Don’t forget to ask “Is it worth continuing?”
- Look for ‘key’ references: Cochrane Risk of Bias, GRADE, PRISMA
- I^2 >50%: adequate statistical heterogeneity to suggest looking deeper into clinical, methodological heterogeneity reported
- Would your patient meet the inclusion criteria of trials/studies in the review?

Publication Bias: Solution

- All trials registered at inception,
  - The National Clinical Trials Registry: Cancer Trials
  - National Institutes of Health Inventory of Clinical Trials and Studies
  - International Registry of Perinatal Trials
- Meta-Registry of trial Registries
  - www.clinicaltrials.org
  - www.controlled-trials.com

COCHRANE & GRADE

THE COCHRANE COLLABORATION®

Quality assessment							Su	
Participants (studies) Follow up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%)	
							With Aspirin vs. Clopidogrel	With Aspirin + Prasugrel
Mortality, All-Cause (CRITICAL OUTCOME)								
13608 (1 study[2]) 15 months	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision[1]	undetected	⊕⊕⊕⊕ HIGH[1]	197/6795 (2.9%)	188/6813 (2.8%)
Mortality, Cardiovascular (CRITICAL OUTCOME)								


JOURNAL OF EVIDENCE-BASED MEDICINE

Journal of Evidence-Based Medicine ISSN 1756-5391

METHODOLOGY

Understanding GRADE: an introduction

Gabrielle Goldet and Jeremy Howick

Department of Primary Health Sciences, University of Oxford, Oxford, UK

Abstract

Objective: Grading of recommendations, assessment, development, and evaluations (GRADE) is arguably the most widely used method for appraising studies to be included in systematic reviews and guidelines. In order to use the GRADE system or know how to interpret it when reading reviews, reading several articles and

Keywords
Evidence-based medicine; GRADE; randomized controlled trial; systematic review.

Correspondence

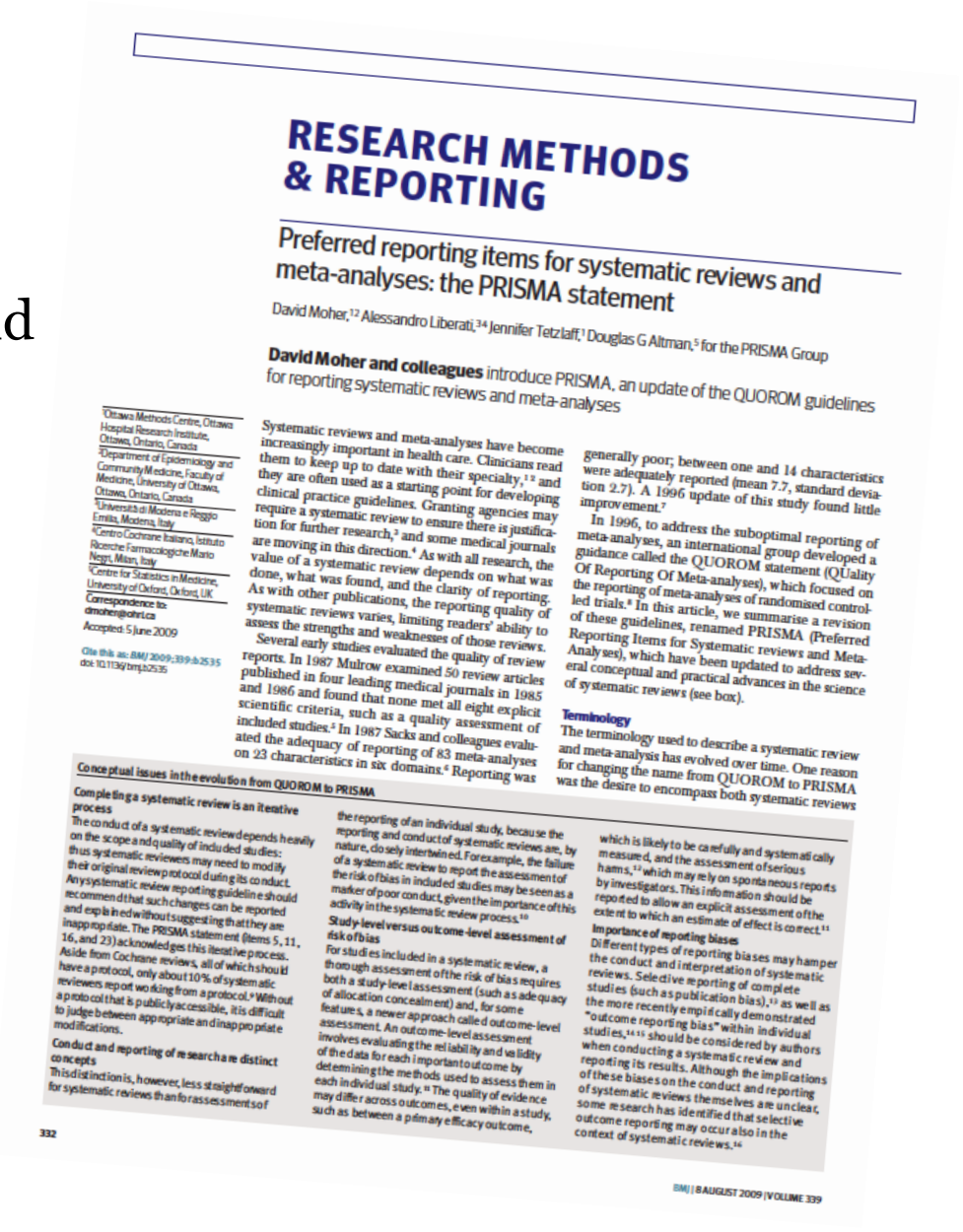
JEBM; 6:50-54

PRISMA (QUORUM)

***P**referred **R**eporting **I**tems for **S**ystematic **R**eviews and **M**eta-**A**nalyses*

- Consists of a 27-item checklist and four phase flow diagram
- Evidence-based minimum set of items for reporting in systematic reviews and meta-analyses
- Can be used for critical appraisal but not designed for it

http://www.prisma-statement.org/



RESEARCH METHODS & REPORTING

Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement

David Moher,[1,2] Alessandro Liberati,[3,4] Jennifer Tetzlaff,[1] Douglas G Altman,[5] for the PRISMA Group

David Moher and colleagues introduce PRISMA, an update of the QUOROM guidelines for reporting systematic reviews and meta-analyses

[1]Ottawa Methods Centre, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada [2]Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada [3]Università di Modena e Reggio Emilia, Modena, Italy [4]Centro Cochrane Italiano, Istituto Ricerche Farmacologiche Mario Negri, Milan, Italy [5]Centre for Statistics in Medicine, University of Oxford, Oxford, UK

Correspondence to: dmoher@ohri.ca

Accepted: 5 June 2009

Cite this as: BMJ 2009;339:b2535 doi:10.1136/bmj.b2535

Systematic reviews and meta-analyses have become increasingly important in health care. Clinicians read them to keep up to date with their specialty,[1 2] and they are often used as a starting point for developing clinical practice guidelines. Granting agencies may require a systematic review to ensure there is justification for further research,[3] and some medical journals are moving in this direction.[4] As with all research, the value of a systematic review depends on what was done, what was found, and the clarity of reporting. As with other publications, the reporting quality of systematic reviews varies, limiting readers' ability to assess the strengths and weaknesses of those reviews.

Several early studies evaluated the quality of review reports. In 1987 Mulrow examined 50 review articles published in four leading medical journals in 1985 and 1986 and found that none met all eight explicit scientific criteria, such as a quality assessment of included studies.[5] In 1987 Sacks and colleagues evaluated the adequacy of reporting of 83 meta-analyses on 23 characteristics in six domains.[6] Reporting was generally poor; between one and 14 characteristics were adequately reported (mean 7.7, standard deviation 2.7). A 1996 update of this study found little improvement.[7]

In 1996, to address the suboptimal reporting of meta-analyses, an international group developed a guidance called the QUOROM statement (QUality Of Reporting Of Meta-analyses), which focused on the reporting of meta-analyses of randomised controlled trials.[8] In this article, we summarise a revision of these guidelines, renamed PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses), which have been updated to address several conceptual and practical advances in the science of systematic reviews (see box).

Terminology

The terminology used to describe a systematic review and meta-analysis has evolved over time. One reason for changing the name from QUOROM to PRISMA was the desire to encompass both systematic reviews

Conceptual issues in the evolution from QUOROM to PRISMA

Completing a systematic review is an iterative process

The conduct of a systematic review depends heavily on the scope and quality of included studies: thus systematic reviewers may need to modify their original review protocol during its conduct. Any systematic review reporting guideline should recommend that such changes can be reported and explained without suggesting that they are inappropriate. The PRISMA statement (items 5, 11, 16, and 23) acknowledges this iterative process. Aside from Cochrane reviews, all of which should have a protocol, only about 10% of systematic reviewers report working from a protocol.[9] Without a protocol that is publicly accessible, it is difficult to judge between appropriate and inappropriate modifications.

Conduct and reporting of research are distinct concepts

This distinction is, however, less straightforward for systematic reviews than for assessment of the reporting of an individual study, because the reporting and conduct of systematic reviews are, by nature, closely intertwined. For example, the failure of a systematic review to report the assessment of the risk of bias in included studies may be seen as a marker of poor conduct, given the importance of this activity in the systematic review process.[10]

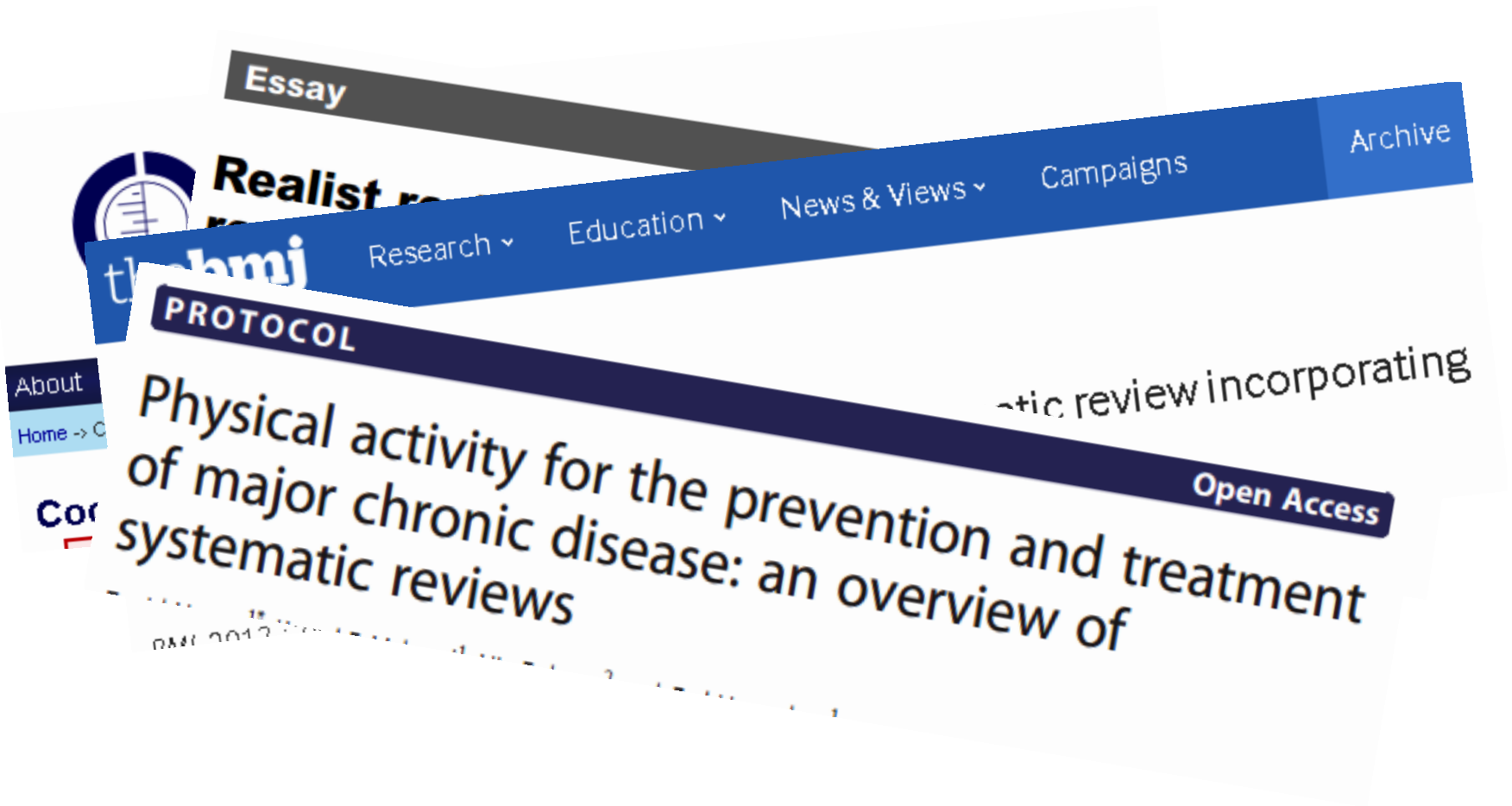
Study level versus outcome level assessment of risk of bias

For studies included in a systematic review, a thorough assessment of the risk of bias requires both a study level assessment (such as adequacy of allocation concealment) and, for some features, a newer approach called outcome-level assessment. An outcome-level assessment involves evaluating the reliability and validity of the data for each important outcome by determining the methods used to assess them in each individual study.[11] The quality of evidence may differ across outcomes, even within a study, such as between a primary efficacy outcome, which is likely to be carefully and systematically measured, and the assessment of serious harms,[12] which may rely on spontaneous reports by investigators. This information should be reported to allow an explicit assessment of the extent to which an estimate of effect is correct.[11]

Importance of reporting biases

Different types of reporting biases may hamper the conduct and interpretation of systematic reviews. Selective reporting of complete studies (such as publication bias),[13] as well as the more recently empirically demonstrated "outcome reporting bias" within individual studies,[14 15] should be considered by authors when conducting a systematic review and reporting its results. Although the implications of these biases on the conduct and reporting of systematic reviews themselves are unclear, some research has identified that selective outcome reporting may occur also in the context of systematic reviews.[16]

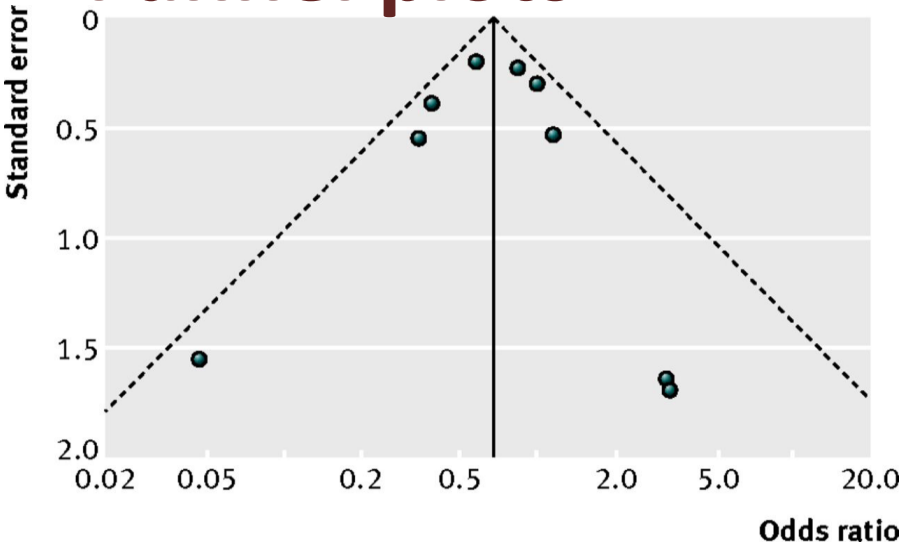
Coming soon….already here?

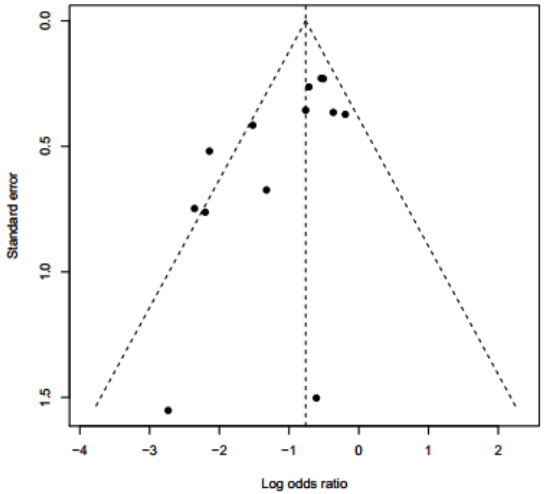


Publication bias

- Occurs when publication of research results depends on their nature and direction
- Often happens because smaller (n and effect size) studies not submitted/rejected, selective reporting, selective citation (of +ve results)
- Funnel plots help identify if there is a bias:
  - Treatment effect vs. study size
  - Smaller the study = wider the effects
  - Largest studies will be near the average (truth), small studies will spread on both sides = symmetric funnel
  - Asymetric funnel indicates publication bias – but not all the time (e.g. heterogeneity)
  - Interpretation difficult if only a few studies in meta-analysis

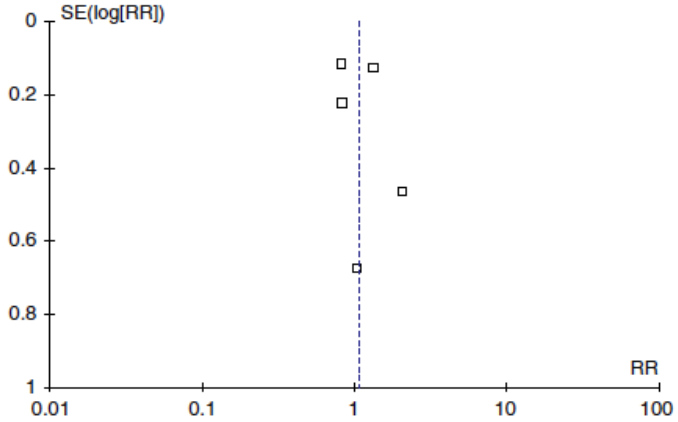
Funnel plots





⇒ hints for publication bias

Knee Surg Sports Traumatol Arthrosc (2010) 18:304–311



RR – Relative risk

SE (log[RR]) – standard error via logarithmic transformation of relative risk

Funnel plot examples

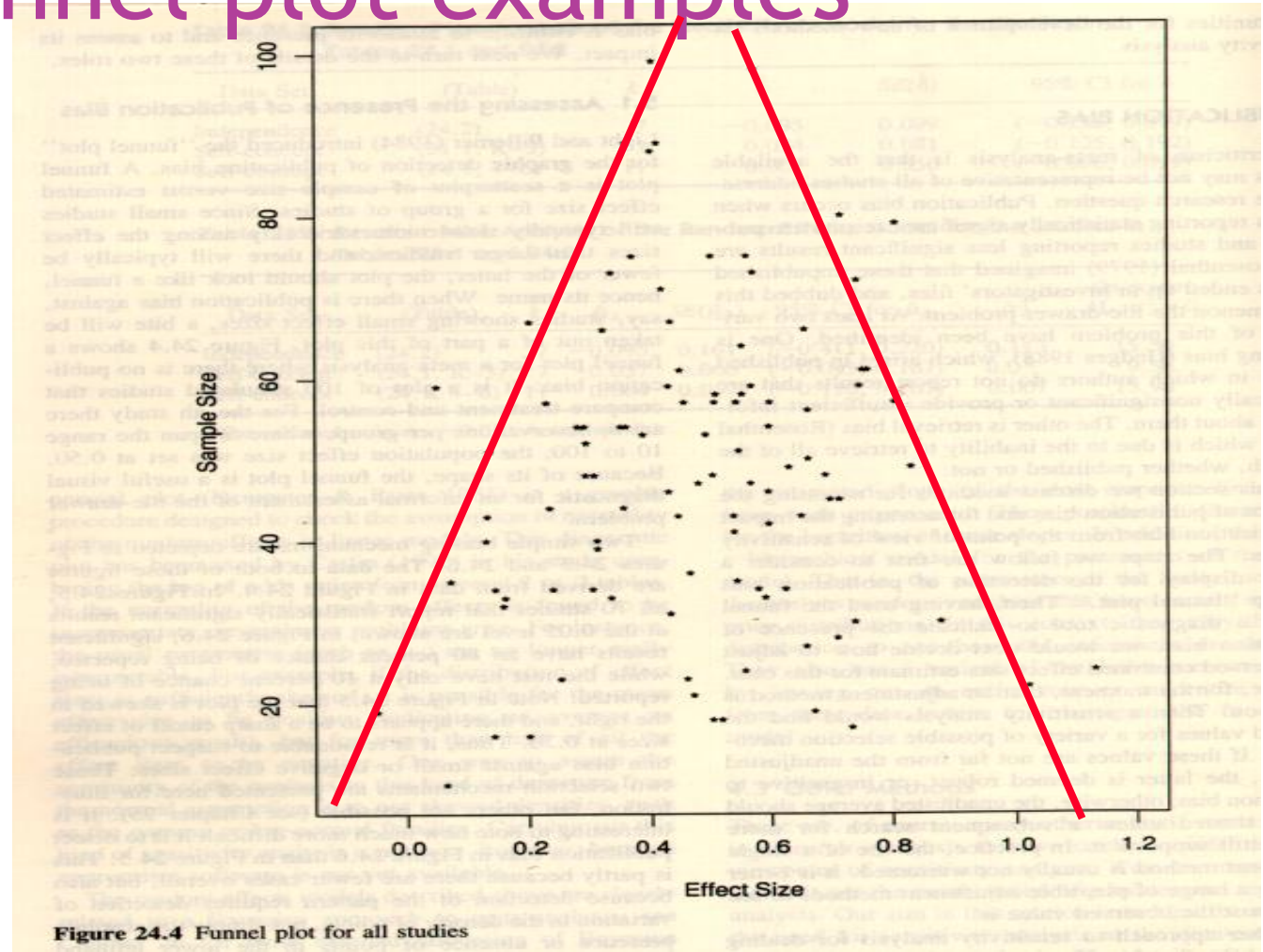
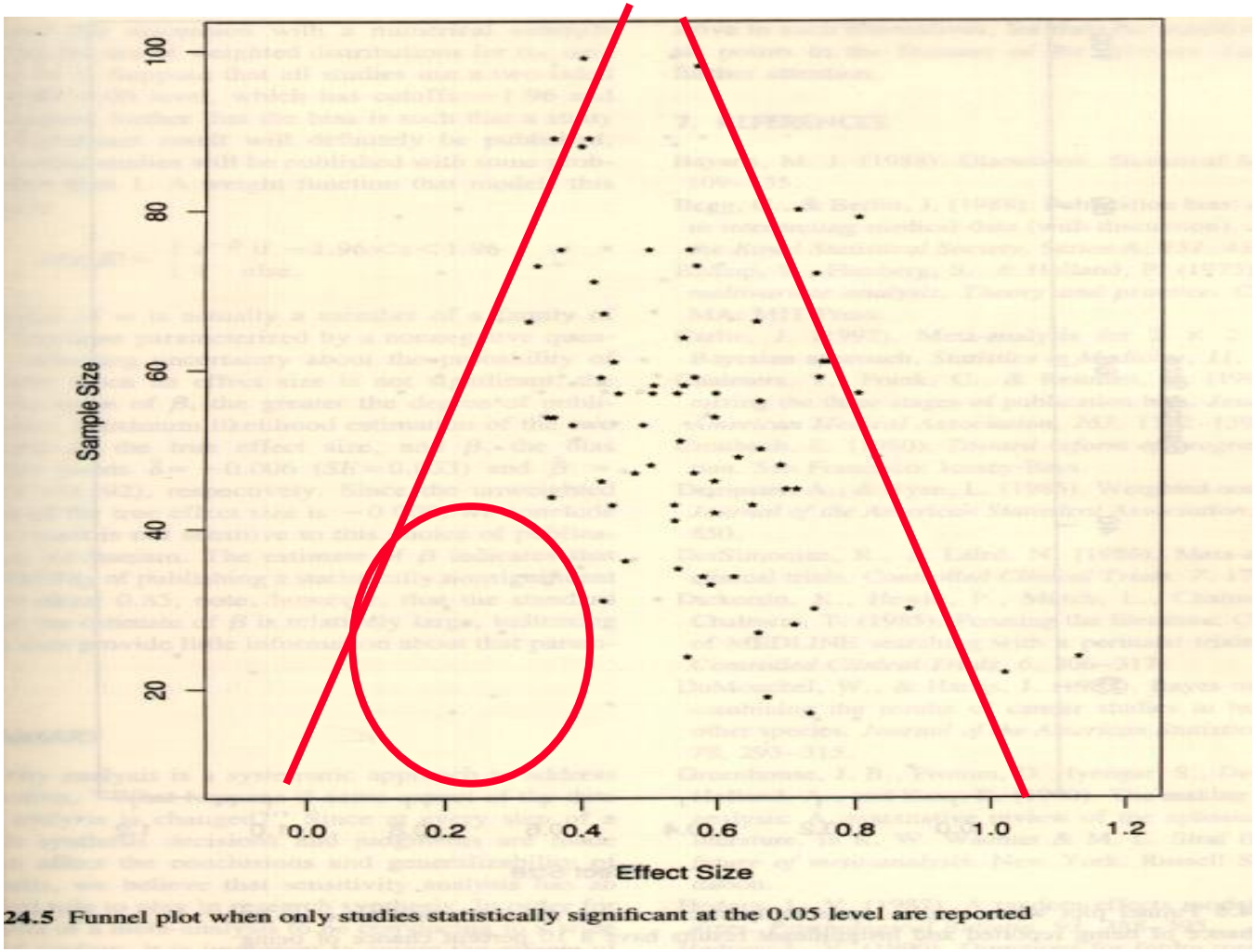


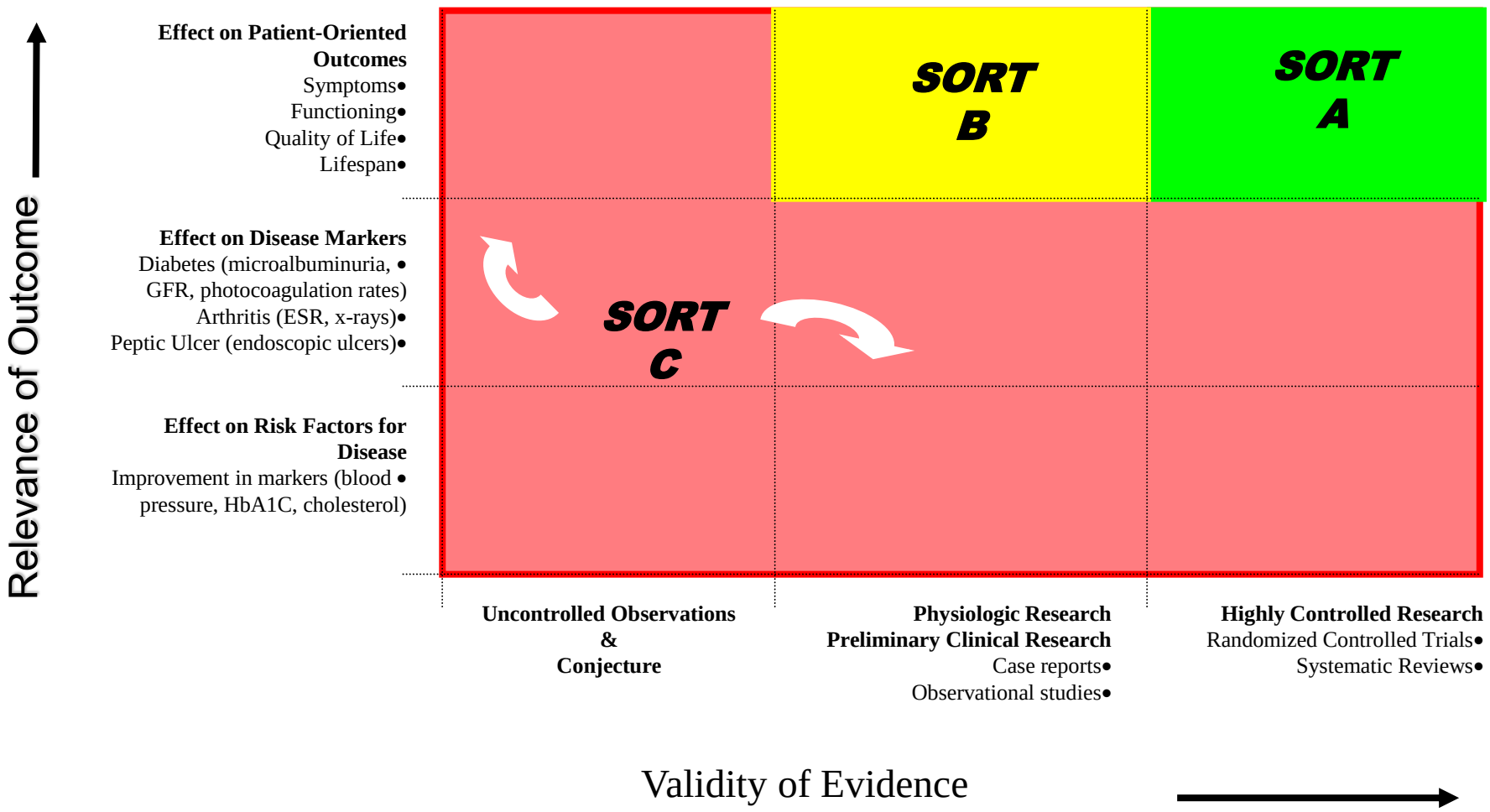
Figure 24.4 Funnel plot for all studies

From: Cooper & Hedges: The handbook of research synthesis. 1994



24.5 Funnel plot when only studies statistically significant at the 0.05 level are reported

From: Cooper & Hedges: The handbook of research synthesis. 1994


Relevance of Outcome
Effect on Patient-Oriented Outcomes
Symptoms•
Functioning•
Quality of Life•
Lifespan•
Effect on Disease Markers
Diabetes (microalbuminuria, •
GFR, photocoagulation rates)
Arthritis (ESR, x-rays)•
Peptic Ulcer (endoscopic ulcers)•
Effect on Risk Factors for Disease
Improvement in markers (blood •
pressure, HbA1C, cholesterol)
SORT B
SORT A
SORT C
Uncontrolled Observations & Conjecture
Physiologic Research
Preliminary Clinical Research
Case reports•
Observational studies•
Highly Controlled Research
Randomized Controlled Trials•
Systematic Reviews•
Validity of Evidence

Practice Guidelines

Practice Guidelines

- The Good,
- The Not-So-Good
- The Ugly

Where do practice guidelines come from?

- **Trust us, we're the experts**: Opinion-based/ consensus guidelines
  - Whose opinion? Do they have a conflict of interest? What is their perspective?
- **Trust us, we have the evidence:** "Evidence-based"
  - How was the evidence used? Patient-oriented? Values?
- **Evidence-linked:**
  - Here is how we found the evidence, used the evidence
  - Strength of recommendation noted

What is a Clinical Practice Guideline?

- A clinical practice guideline is a systematically developed statement designed to help health care professionals make decisions about appropriate health care for specific clinical circumstances.

Guidelines attempt to do this by:

- Describing a range of generally accepted approaches for the diagnosis, management, or prevention of specific diseases or conditions.

- Defining practices that meet the needs of most patients in most circumstances.

Questions to Ask

Scope and Purpose

- Are the overall objectives of the guideline specifically described?
- Is the clinical question(s) covered by the guideline specifically described?
- Are the patients to whom the guideline is meant to apply specifically described?

Stakeholder Involvement

- Does the guideline development group include individuals from all of the relevent professional groups?
- Have the patients' views and preferences been sought?
- Have the target users of the guideline been clearly defined?
- Has the guideline been piloted among target users?

Rigour of Development

- Were systematic methods used to search for evidence?
- Is the criteria for selecting the evidence clearly described?
- Were the methods used for formulating the recommendations clearly described?
- Were the health benefits, side effects, and risks considered in formulating the recommendations?
- Is there an explicit link between the recommendations and the supporting evidence?
- Has the guideline been externally reviewed by experts prior to its publication?
- Is a procedure for updating the guideline provided?

Clarity and Presentation

- Are the recommendations specific and unambiguous?
- Are different options for management of the condition clearly presented?
- Are key recommendations clearly identifiable?
- Is the guideline supported with tools for application?

Applicability

- Have the potential organization barriers in applying the recommendation been discussed?
- Have the potential cost implications of applying the recommendations been considered?
- Does the guideline present key review criteria for monitoring and/or auditing purposes?

Editorial Independence

- Is the guideline editorially independent from the funding body?
- Have conflicts of interest of guideline development members been recorded?

Appraisal of Clinical Practice Guidelines

The AGREE

- The AGREE (Appraisal of Guidelines for Research & Evaluation) Instrument is a tool that assesses the methodological rigour and transparency in which a guideline is developed and it is used internationally.
- The purpose of the AGREE Instrument is to provide a framework for assessing the quality of clinical practice guidelines.

AGREE

Advancing the science of practice guidelines

Login

Search Go

Home | About | AGREE Tools | Research Projects | News | My AGREE PLUS

- AGREE II
- My AGREE PLUS
- AGREE GRS Instrument
- AGREE-REX: Recommendation EXcellence
- AGREE-HS: Health Systems
- AGREE Reporting Checklist
- CheckUp
- **Original AGREE Instrument**
  - Original AGREE Instrument Publications
  - Original AGREE Instrument Translations
- Guideline Implementability for

Original AGREE Instrument

Since its original release in 2003, the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument has advanced the science of practice guidelines (PG) appraisal and quickly became the international gold standard for PG evaluation and development.

The AGREE Instrument quickly became the international gold standard for practice guidelines evaluation and development.

- **Read more about the history of AGREE**

The AGREE Instrument has been **translated into many languages**, endorsed by several organizations (e.g., National Institute for Health and Clinical Excellence), and used by many development groups (e.g., WHO Advisory Committee on Health Research).

A page of key publications of the original AGREE instrument can be accessed **here.**

We recommend the use of the new AGREE II in place of the original AGREE Instrument.

AGREE II Instrument

Download the AGREE II

Training tools

Learn how to apply the AGREE II though our training modules.

Appraise guidelines

Appraise practice guidelines with the My AGREE PLUS online appraisal platform.

Guideline Reporting

Apply the AGREE Reporting Checklist when reporting guidelines.

Follow us on Twitter

Follow @AGREEScientific

The AGREE

- The AGREE Instrument is designed to assess guidelines developed by local, regional, national, or international groups or affiliated government organizations. These include:
  - New guidelines
  - Existing guidelines
  - Updates of existing guidelines

http://www.agreetrust.org

The AGREE II

- The original AGREE Instrument has been updated and methodologically refined. The AGREE II is now the new international tool for the assessment of practice guidelines. The AGREE II is both valid and reliable and comprises 23 items organized into the original 6 quality domains.

Find Clinical Practice Guidelines:

- Search the National Guideline Clearinghouse
- Go to ClinicalKey and click on the link to "Practice Guidelines" at the top of the page.
- When searching MEDLINE, click on "Additional Limits", and limit by publication type (Practice Guideline).

National Guideline Clearinghouse

- www.ngc.gov
- Vetted guidelines from various groups
- Standard organization so that information can be compared across various guidelines

References

- Critical Appraisal tools and Work sheets http://www.cebm.net/critical-appraisal/
- Jane M Young & Michael J Solomo. How to critically appraise an article. Nature Reviews Gastroenterology and Hepatology 6, 82-91 (February 2009) | doi:10.1038/ncpgasthep1331
- What is a critical appraisal? http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what_is_critical_appraisal.pdf
- The BMJ - How to Read a Paper http://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper
- Evidently Cochrane http://www.evidentlycochrane.net/
- CEBM http://www.cebm.net/

References

- http://www.casp-uk.net/e-learning
- http://www.casp-uk.net/checklists
- http://clinicalevidence.bmj.com/x/set/static/ebm/toolbox/665061.html
- Critical Appraisal tools and Work sheets http://www.cebm.net/critical-appraisal/
- Jane M Young & Michael J Solomo. How to critically appraise an article. Nature Reviews Gastroenterology and Hepatology 6, 82-91 (February 2009) | doi:10.1038/ncpgasthep1331
- What is a critical appraisal? http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what_is_critical_appraisal.pdf
- The BMJ - How to Read a Paper http://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper
- Evidently Cochrane http://www.evidentlycochrane.net/
- CEBM http://www.cebm.net/