Basics of Critical Appraisal of Available Evidence - I

(Therapy, Diagnosis and Prognosis)

Dr. Shatha Alduraywish, MBBS; MSc; PhD

Assistant Professor

Department of Family and Community Medicine

College of Medicine, KKUH

King Saud University

Dr. Nada A. AlYousefi

Assistant Professor, Postgraduate Trainer and Consultant

Family Medicine Unit

Department of Family and Community Medicine

College of Medicine(KSU)

February 2018

Session Overview

- Introduction
- Therapy Article Critique
- Diagnosis Article Critique
- SR and MA

Structure of an article

- Title
- Abstract
- Introduction
- Background / review of literature
- Organizational context
- Methodology
- Results
- Discussion

Structure of an article

1. Title
   - Should be informative
2. Abstract
   - contains a summary of the the research question, key methods, results and conclusions of the study
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. Organizational context (Research setting)

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

Structure of an article

8. Discussion

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

What is Critical appraisal

- The process of **carefully and systematically** examining research to judge its trustworthiness, and its value and relevance in a particular context.

- It is an **essential skill for evidence-based practice** because it allows public health professionals and clinicians to find and use research evidence reliably and efficiently to inform their decision-making.

General Tips

Any study should have the following criteria to be relevant:

- The topic addresses a common problem in my practice
- Patient oriented evidence that matter (POEM) outcomes like improvement of symptoms, quality of life, cost... etc.
- Results (if valid) will change my practice.

In general, for critical appraisal, look for:

- Validity
- Results
- Applicability.

Appraising The Evidence

1. Is the study valid?
   - Evaluating its methodological quality
   - Decide whether studies have been undertaken in a way that makes their findings reliable.
2. What are the results?
   - Whether the study's results are clinically important)
   - Make sense of the results.
3. Are the results useful?
   - How the study results applies to your question
   - Know what these results mean in the context of the decision that needs to be made.

IMPORTANT SAFETY INFORMATION

Serious risks associated with TOPAMAX include lowered bicarbonate levels in the blood resulting in an increase in the acidity of the blood (metabolic acidosis), and hyperventilation (rapid, deep breathing) or fatigue. More severe symptoms of metabolic acidosis could include irregular heartbeat or changes in the level of alertness. Chronic, untreated metabolic acidosis may increase the risk for kidney stones or bone disease. Your doctor may want to do simple blood tests to measure bicarbonate levels.

ORTHO-McNEIL NEUROLOGICS, ©OMN, Inc. 2006 November 2006 02M749AR1

Do you *worry* about migraines even when you're not having one?

You know the routine. First comes the pain.
Then you take a medication to treat the pain.
Then you worry about when the next migraine might strike.

And that anticipation can be enough to distract you from the things you like to do. It can feel like you're trapped in a cycle of suffering, treating and worrying. But there is something you can do.

TOPAMAX can help change your migraine cycle.

While migraines can't be completely eliminated, TOPAMAX helps stop them before they start, so you can get fewer of them to worry about.

Unlike treatments you use at the start of a migraine, TOPAMAX works differently. It's a daily prescription medication you take for as long as you and your healthcare professional decide you need it.

So ask your healthcare professional about helping to change your migraine cycle today. It could be the change you're looking for.

TOPAMAX is approved for migraine prevention in adults only. TOPAMAX is not for the acute treatment of migraines.

TOPAMAX may cause side effects, so talk to your healthcare professional to see if it could be an option for you. Please see Important Safety Information below.

Life shouldn't always revolve around migraines.

TOPAMAX®
(topiramate) Tablets

www.TOPAMAX.com

Other serious risks include increased eye pressure (glaucoma), decreased sweating, increased body temperature, kidney stones, sleepiness, dizziness, confusion, and difficulty concentrating. Tell your doctor immediately if you have blurred vision or eye pain.

More common side effects are tingling in arms and legs, loss of appetite, nausea, diarrhea, taste change and weight loss.

Tell your doctor about other medications you take.

Please see accompanying important information about TOPAMAX on adjacent page.

Together Rx Access

Together Rx Access is a trademark of Together Rx Access, LLC.

Clarification

Now let's say that the same patient has heard from a friend that there is a vitamin that will help prevent migraines. What study design could answer the question of whether there is a vitamin that is useful in preventing migraine headaches in this patient?

Study Methods to Answer This Question

- Epidemiology: Patients taking a vitamin are less likely to have migraines
- Pharmacology: Drug x affects cerebral vasculature in rat brain isolates
- Case report: “It worked on one patient”
- Case-series: “It worked on a bunch of patients”
- Randomized controlled trial: 1/2 get drug, 1/2 placebo. No one knows who ‘til the end who took what

Therapy Article Critique

The best for therapy evidence is Systemic Review article or Meta-analysis of **Randomized Controlled Trial (RCT).** If you didn't find these articles, then RCT will be fine.

Clinical Trial Compares

INTERVENTION

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

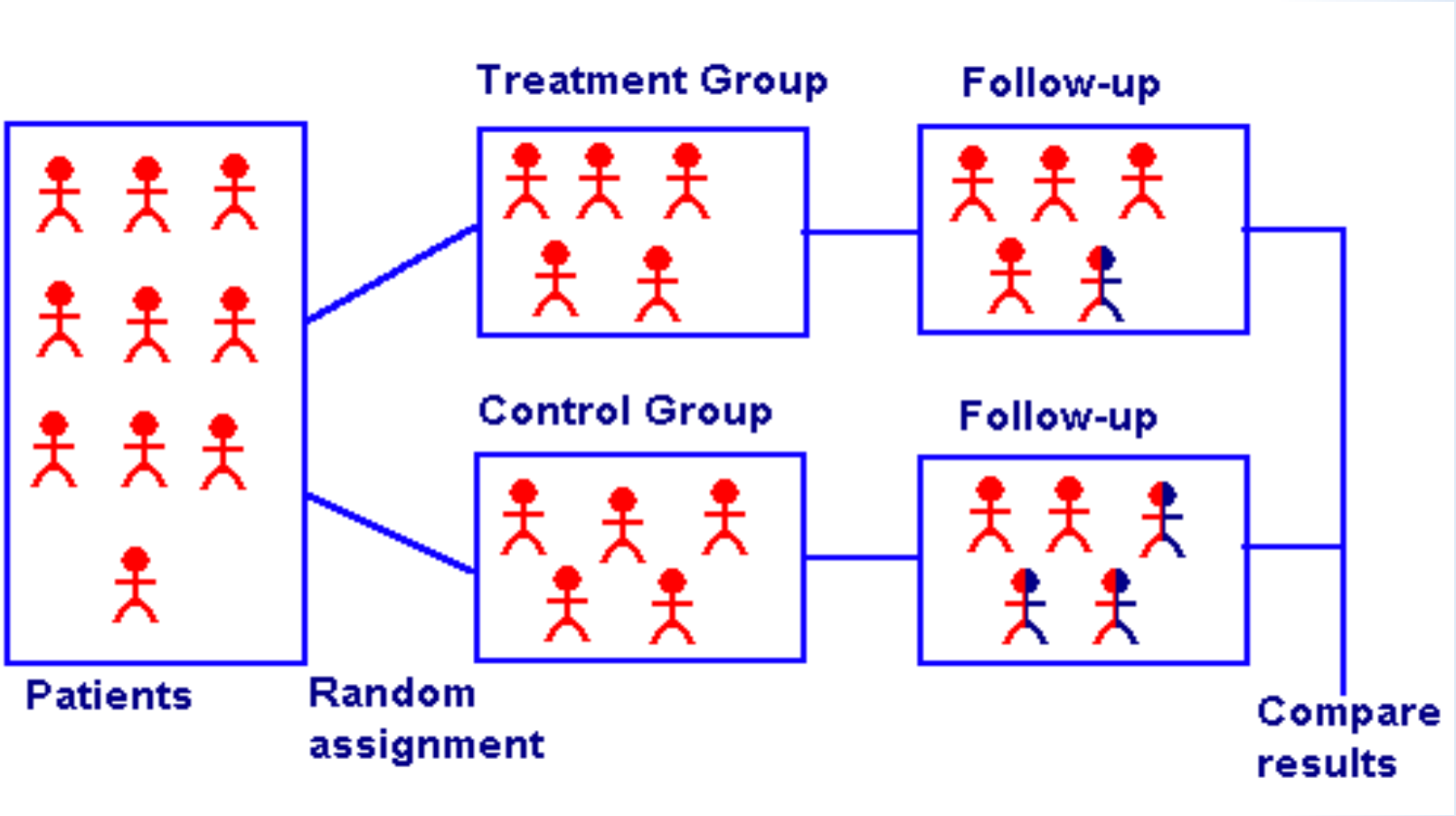
CONTROL

- 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

Randomized Controlled Trial (**RCT**)



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

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.

Validity

Validity Mnemonics (RABI)

- **R** Randomization
- **A** Allocation Concealed assignment. Like opaque envelope or central call allocated.

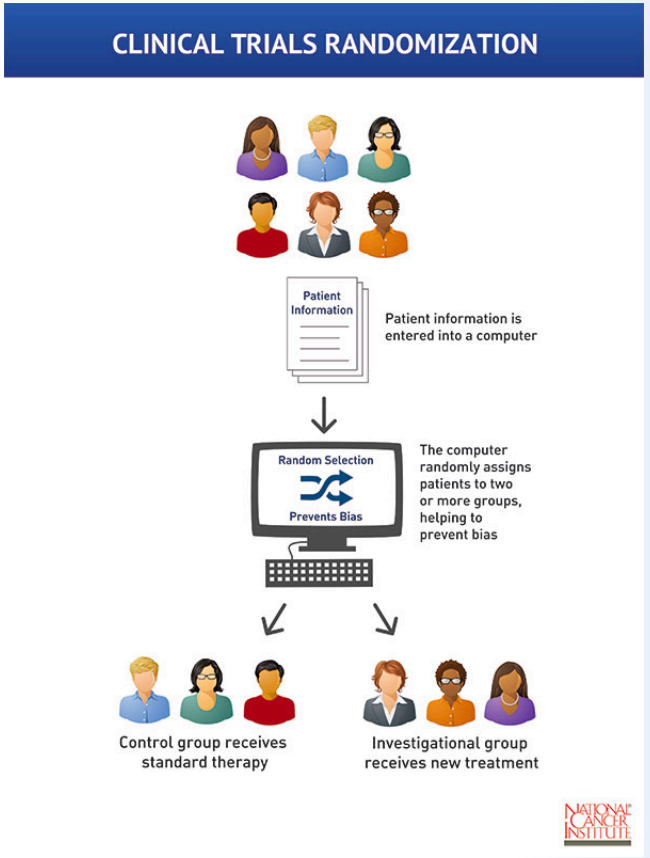
  Attrition (complete follow up)
- **B** Blindness (single, double, triple) 3Cs (contamination, co-intervention & compliance)
- **I** Intention to Treat analysis

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

Question:
Was the assignment of patient randomized?



What is Randomization

- It is a process where each patient has equal chance to be in control group or experiment group (i.e. every one in the sample has 50% chance to be in either group, experiment group or control group).
- It can be done by:
  - **tossing a coin** to randomly allocate the participants.
  - using computer random number generation
  - In blocks.

- Once you can predict where the patient will go (to experiment or control group), then you broke the randomization.

Why Randomization

1. To avoid **confounding factors**; which affect both cause and outcome.
2. To ensure **equal base-line characteristics** in both groups.
3. Both groups are **equal** in known or unknown prognostic factors.

Randomization is done before allocation.

Validity

Validity Mnemonics (RABI)

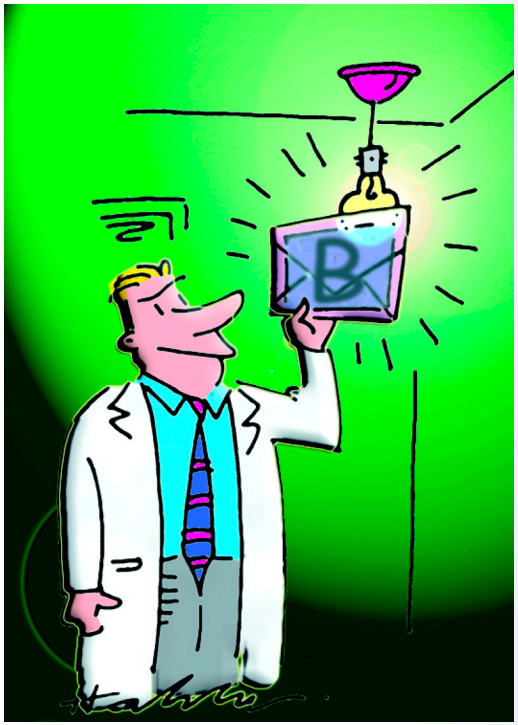
- **R** Randomization
- **A** Allocation Concealed assignment. Like opaque envelope or central call allocated.

  Attrition (complete follow up)
- **B** Blindness (single, double, triple) 3Cs (contamination, co-intervention & compliance)
- **I** Intention to Treat analysis

Allocation (assignment - concealment)

It is allocation of patient to control or experiment group.

Hide the patient..!!



Allocation (assignment - concealment)

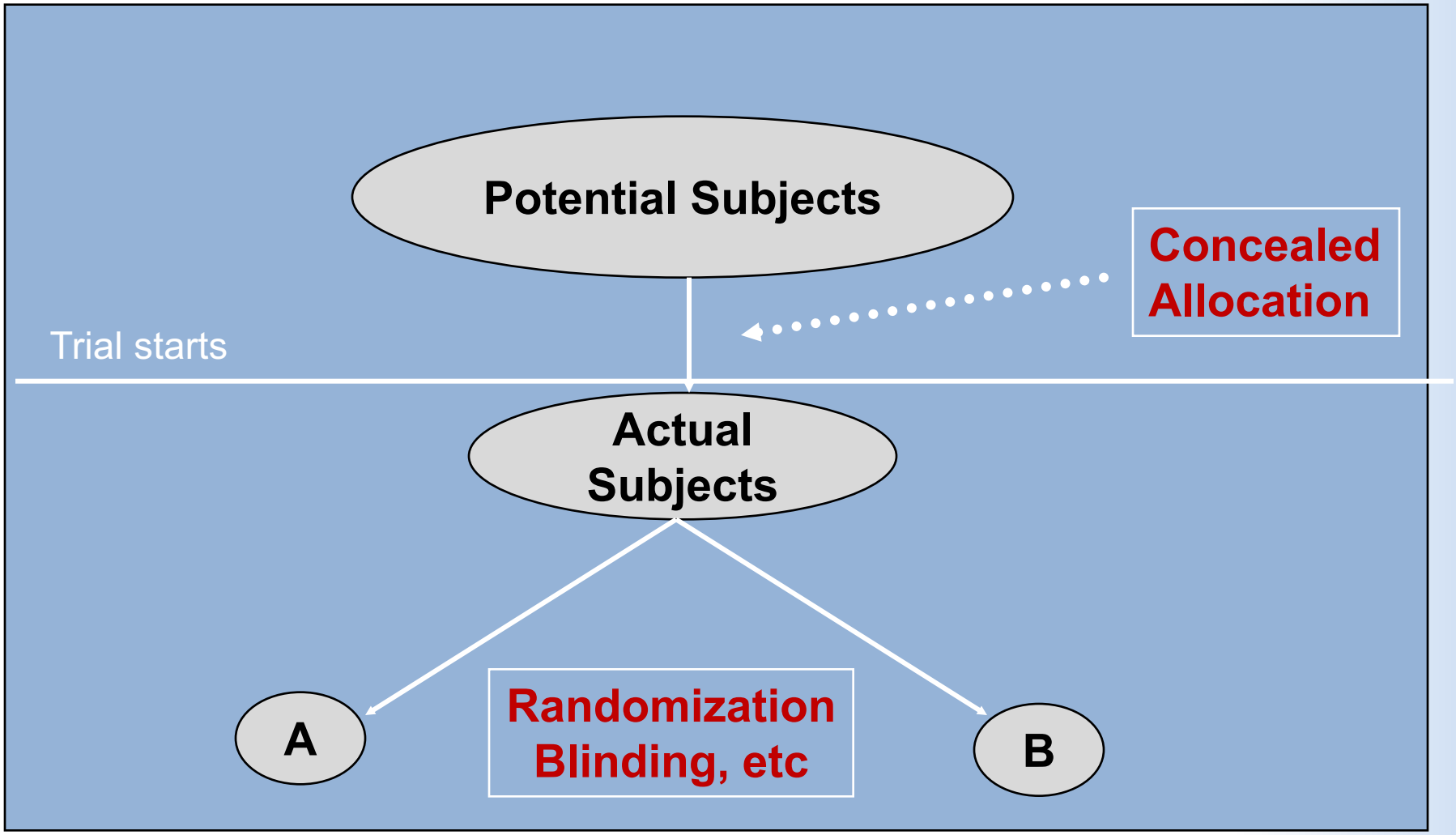
- Allocation comes **after** randomization.
- It is hiding allocation **before treatment** starts and it is meant to prevent **selection bias**.
- Allocation should be **concealed**
  - it is to conceal allocation of study group assignment from those responsible for assessment of patient for entry of trial; i.e. no one from research team knows which patient from which group he is allocated.
- It is done by
  - opaque envelop
  - computerized protected folder

Why allocation concealment?

To maintain randomization.

Trials with unconcealed allocation consistently overestimate benefit by ~40%

Conducting RCT study



Validity

Validity Mnemonics (RABI)

- **R** Randomization
- **A** Allocation Concealed assignment. Like opaque envelope or central call allocated.

  Attrition (complete follow up)
- **B** Blindness (single, double, triple) 3Cs (contamination, co-intervention & compliance)
- **I** Intention to Treat analysis

Follow up
&
Intention To Treat (ITT)

Question:

Were all patient who entered the trial properly accounted for and attributed at its conclusion?

Follow up completion

- Follow up means to follow patient from the time of inclusion in the study until occurrence of primary outcome.
- Accepted when ≥ 80% of patients completed the study (maintain power)
- *Duration of study* should be sufficient for outcome to occur (except in some ethical issues in which the study will be terminated once outcome occurs before finishing the study period)

Types of Analysis in RCT

Intention To Treat (ITT) analysis

- The patients are analyzed in the same group to which they were randomized.
- The number of analyzed patients is the **same** number of patients the trial started with
- **Advantage**:
  - to maintain randomization.
- **Disadvantage**:
  - it gives false effect estimate of experiment or control treatment because of:
    - 1) Including dropped out patients.
    - 2) Including non-compliant patients.
    - 3) Including patients with co-intervention.

Per protocol analysis

- Analyze data from patients who completed the trial **ONLY** and analyzed in the arm in which they finished the trial
- The number of analyzed patients is **lesser** than the number of patients the trial started with
  - because of exclusion "non-compliance, contamination, or co-intervention")
- **Disadvantage**
  - It can not maintain the randomization..!!

Validity

Validity Mnemonics (RABI)

- **R** Randomization
- **A** Allocation Concealed assignment. Like opaque envelope or central call allocated.

  Attrition (complete follow up)
- **B** Blindness (single, double, triple) 3Cs (contamination, co-intervention & compliance)
- **I** Intention to Treat analysis

Blindness

Question:
Were patients, physicians and those during assessment "blind" to treatment?

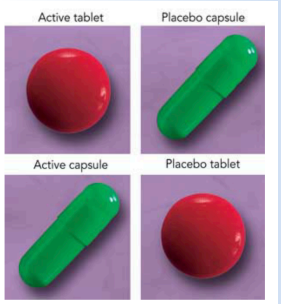
What is Blindness?

- Hide the treatment..!!
  - It is related to treatment; no relation with patient
- It is hiding allocation after treatment starts and it is meant to prevent performance bias.

Blindness could be:

1. Single (either patient or physician blinds to treatment)
2. Double (patient and physician blind to treatment)
3. Triple (patient, physician and data analyst blind to treatment)

How is blindness done?



1. Same shape, color, taste for both control and experiment treatment.
2. Dummy tablets
    1. Experiment group → active intervention + dummy control
    2. Control group → active control + dummy intervention

Note

In some studies you can't keep blindness, e.g. surgical intervention

Why blindness?

1. To maintain randomization.
2. To minimize contamination
   - Any member from one group received treatment from other group that is included in the study
3. To minimize co-intervention
   - Any extra intervention other than study treatment to either group; like outside drug

Five groups can be blind:

patients, clinicians, outcome assessors, data collectors, and data analysts.

Question: Were both groups similar at start of trial?

Equal percentages of:

- Demographics data
- Co-morbidities
- Severity
- Confounding factors
- Prognostic factors

Sometimes even with randomization, both groups are not similar

(need adjustment e.g. by logistic regression… etc to remove the effect of remaining confounding factors).

Question: Aside from experiment, were both groups treated equally?

Contamination

- Any member from one group received treatment from other group that is **included in the study**.
- This could be non-pharmacological contamination like when the control group has adapted other behavior from experiment group like exercise.

Co-intervention

- Any **extra intervention** other than study treatments to either group, like outside drug.

Compliance

- Was it mentioned?
- How is it looked for?

Types of bias in RCT

- **Selection bias:**
  - If randomization was not proper or no concealment.
- **Attrition:** lost to follow up.
- **Performance:**
  - If no blindness, there is contamination, co-intervention, or placebo effect
- **Detection (measurement):**
  - When the outcome assessor is not blind during measuring the outcome.

Results

Questions:

1.How large was the treatment effect?

2.How precise was the estimate of the treatment effect?

Question: How large was the treatment effect?

- RR, ARR, RRR, and NNT

Relative Risk (RR) or Risk Ratio:

- $RR = \frac{\text{EER (Experiment Event Rate)}}{\text{CER (Control Event Rate)}}$

	Disease (cases)	No Disease (control)	Row Totals
Exposure (treatment)	a	b	a + b
No Exposure (no treatment)	c	d	c + d
Column Totals	a + c	b + d	a + b + c + d

EER= a/a+b
CER= c/c+d

Relative Risk (RR)

- It could be > 1, or < 1 or equal 1 (no effect)
- If the outcome is **harm**:
  - **> 1** means the experiment intervention is **causing more harm** compared to control.
  - **< 1** means the experiment intervention is **causing less harm** compared to control.
- If the outcome is **benefit**
  - **> 1** means the experiment intervention is **causing more benefit** compared to control.
  - **< 1** means the experiment intervention is **causing less benefit** compared to control.

Interpretation

Risk of having the outcome in experiment group is (x) times the risk in control group.

- RR doesn't tell you the magnitude of benefit of treatment. It only tells *there is increase or decrease risk* in experiment group compared to control group.

Absolute Risk Reduction (ARR)

- Risk Difference.
- ARR = CER – EER
- It tells the **magnitude of benefit**

Interpretation

if 100 patients were treated with experiment treatment, (x) cases of outcome can be prevented.

- **Example:** if **ARR = 15%** in comparing ACEI vs placebo in decreasing IHD. This means **if 100 patients were treated with ACEI, 15 cases of IHD can be prevented compared to placebo.**

- If ARR equals **0**, then there is no difference between experiment and control.

Number Needed to Treat (NNT)

- Number of persons who would have to receive an intervention for 1 to benefit.

NNT=1/ARR

Relative Risk Reduction (RRR)

- RRR = 1 – RR
- It tells how much the experiment treatment is reducing the chance of having outcome in single treated patient.

Interpretation:

- Using experiment treatment will relatively reduce the risk of having the outcome by (%) compared to control treatment.
- **Example**: if **RRR = 70%** in comparing ACEI vs placebo in decreasing IHD.
  - This means **treatment with ACEI will relatively reduce the risk of having IHD by 70% compared to placebo**; i.e.
  - in person using ACEI, his chance of having IHD will be reduced by 70%

Result Tabulation

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

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

Calculations

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

Question: How precise was the estimate of the treatment effect?

- Look at **confident interval (CI)**
- The narrower CI, the more precise.

Precision has nothing to do with statistical significant (p-value)

- i.e. you could have very narrow and precise CI that crosses 1 (line of no effect) which is not statistically significant, or vise versa..!!

Applicability

Question: ***Can the results be applied to my patient?***

Question: ***Were all clinically important outcomes considered?***

- You should see what the means are used by the researcher to reach into outcome.
- Did he used direct method or surrogate (indirect – substitute) outcome *like using lipid profile as an indirect measure of occurrence of IHD*.
- Were outcomes Disease Oriented Evidence (DOE) or Patient-Oriented Evidence that Matters (POEM).
  - *POEM is better like improvement of symptoms, improvement of quality of life… etc.*
- Treatment availability, and is it affordable by the patient (cost benefit).

Question: ***Are the likely treatment benefits greater than the potential harms and costs?***

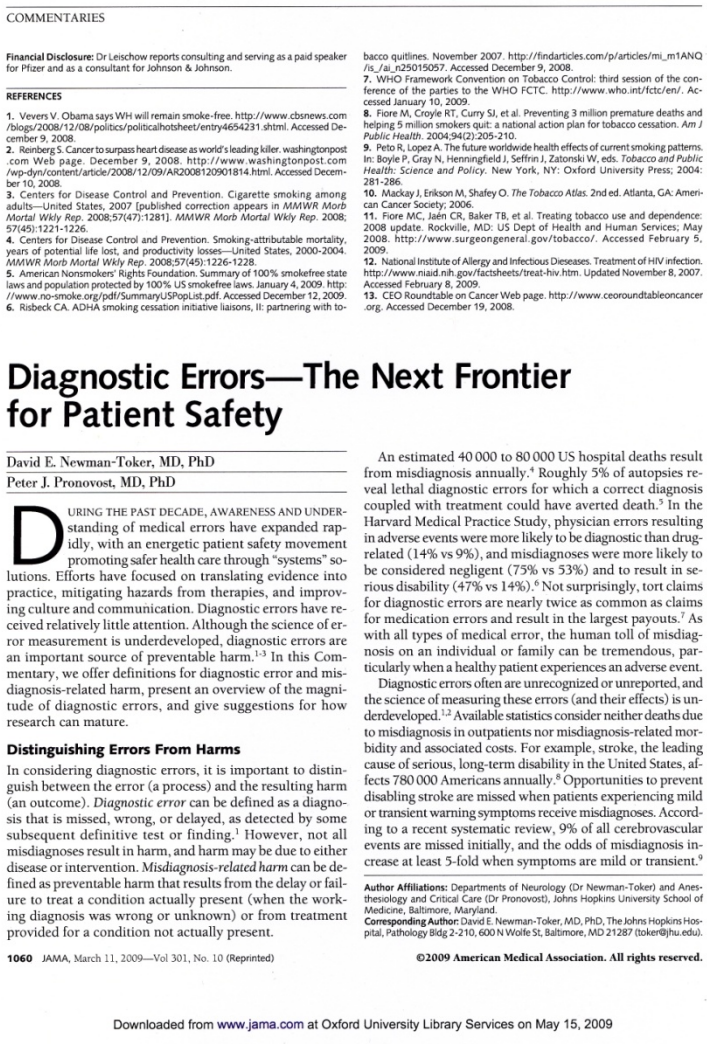
- Weight risk and benefit

Summary

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

Diagnosis Article Critique

The best study type for diagnosis study is ***prospective blind comparison to a gold standard cross-sectional study.***



COMMENTARIES

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).

©2009 American Medical Association. All rights reserved.

Downloaded from www.jama.com at Oxford University Library Services on May 15, 2009

- 2/3 malpractice claims against GPs in UK

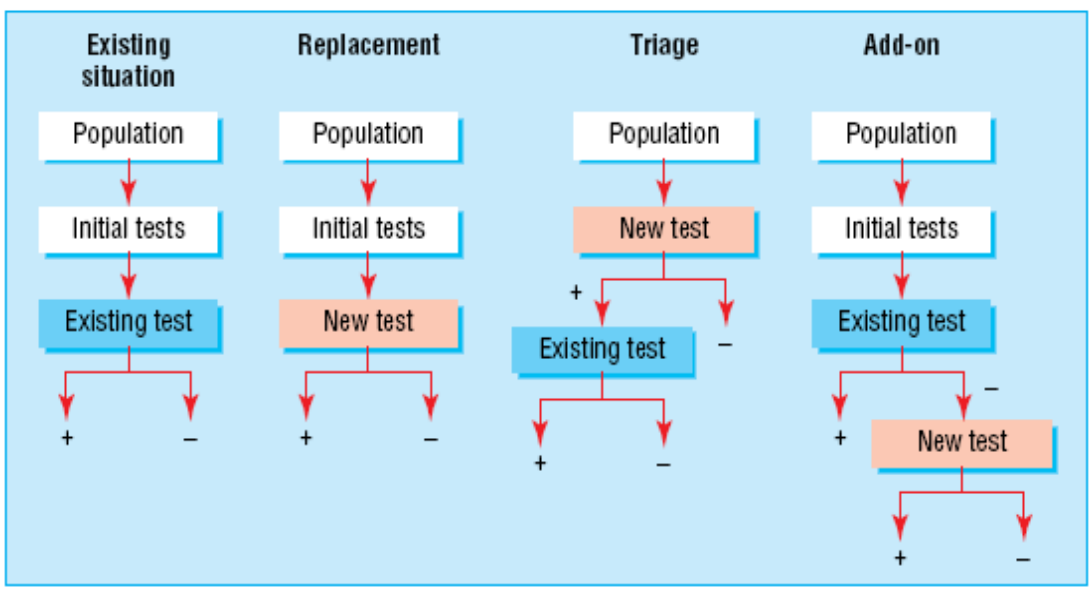
- 40,000-80,000 US hospital deaths from misdiagnosis per year

- 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

Appraising a diagnostic test study using a critical appraisal checklist

Ref. *Mahilum-Tapay L, et al. New point of care Chlamydia Rapid Test – bridging the gap between diagnosis and treatment: performance evaluation study. BMJ 2007;335:1190.*

Why are we looking at the test?

The problem:

An 18-year-old women comes to your clinic because she has pain when passing urine and has noticed a change in her vaginal discharge. You suspect that she has might have *Chlamydia*, but the women hates the idea of going to the hospital or being examined by a clinician and asks if there is a test she can do herself instead

So, we research **alternative test methods for *Chlamydia***

Results of our search

- We find this reference, which assesses a new *Chlamydia* Rapid Test:

  Mahilum-Tapay L, et al. New point of care Chlamydia Rapid Test – bridging the gap between diagnosis and treatment: performance evaluation study. BMJ 2007;335:1190

- The **objective** of this study is to evaluate the performance of a new *Chlamydia* Rapid Test with vaginal swab specimens as a potential tool for *Chlamydia* diagnosis and screening compared with nucleic acid amplification tests with first void urine, and vulvo-vaginal swab specimens.
- Importantly for us, the study also assessed if there is any difference between results of the *Chlamydia* Rapid Test when the swabs are self collected compared with clinician collected

Critical Appraisal

Now we have found a study that may give a solution to our current problem, we need to assess the quality of the research we have found in terms of validity and the importance of the results to see if we can apply this test to the patient.

To do this we can use the critical appraisal **checklist** to evaluate the study

Series of patients

↓

Index test

↓

Reference (“gold”) standard

↓

Compare the results of the index test with the reference standard, blinded

Is the study valid? Screening

Was there a clear question for the study to address?

(the population, test, setting, and outcome)

In this case yes, the study asked:

1. "What is the diagnostic accuracy of the *Chlamydia* Rapid Test compared with polymerase chain reaction and strand displacement amplification assays in the diagnosis of *Chlamydia* in women presenting to a sexual health centre (site 1) and genitourinary medicine clinics (site 2 and 3)?"
2. "Is there a difference in the diagnostic accuracy of the *Chlamydia* Rapid Test between self-collected samples and clinician-collected samples?"

This information can usually be found in the abstract or the introduction to the study

Defining the clinical question: PICO or PIRT

- Patient/Problem
  - How would I describe a group of patients similar to mine?
- Index test
  - Which test am I considering?
- Comparator… or …Reference Standard
  - What is the best reference (gold) standard to diagnose the target condition?
- Outcome….or….Target condition
  - Which condition do I want to rule in or rule out?

Is the study valid? Screening

Is there comparison with an appropriate (gold) reference standard for diagnosing the disorder under assessment?

- ***The reference standard comparison should be the best available indicator of the target disorder***

In this case yes, the study stated that:

> "We assessed the performance of the Chlamydia Rapid Test in order to meet the requirements for Conformité Européenne licensure, which stipulate that the comparator test should be a "state of the art" assay and use specimens approved for the test. Participants from site 1 did not provide endocervical swabs, preventing the pooling of data from all three sites. Given this condition, we chose polymerase chain reaction testing, which is licensed for both urine and endocervical specimens, as the "gold standard" for the study. Studies of Chlamydia trachomatis polymerase chain reaction testing have shown equal performance with cervical and urine specimens, across all volumes of urine tested (<20-90 ml), and good reproducibility. For the genitourinary medicine clinics, endocervical specimens were additionally collected by the clinician and were tested by strand displacement amplification assay at the hospital laboratory."

As the answer is yes to both of our initial screening questions, we should continue with our analysis of the diagnostic test study

Is the study valid? Population

- **Did the study include people with <u>all the common presentations of the target disorder</u>?**

For example, symptoms of early manifestations as well as people with more severe symptoms, and/or people with other disorders that are commonly confused with the target disorder when diagnosing?

Is the study valid? Population

Did the study include people with all the common presentations of the target disorder? For example, symptoms of early manifestations as well as people with more severe symptoms, and/or people with other disorders that are commonly confused with the target disorder when diagnosing?

Yes, the study states that:

> "All women 16 years and over presenting to any of the three sites were invited to participate in the study . Most participants at site 1 were asymptomatic [663 women], in contrast with 441/662 [67%] of the participants from the genitourinary medicine clinics presented with symptoms that included vaginal discharge 305/662 [46%], and lower abdominal pain 149/657 [23%]. In addition 23/668 [3%] of women were diagnosed as having pelvic inflammatory disease."

Is the study valid? Blinding

- Were the people assessing the results of the index diagnostic test blinded to the results of the reference standard?

Is the study valid? Blinding

Were the people assessing the results of the index diagnostic test blinded to the results of the reference standard?

Yes, while the study does not explicitly state blinding, it is very specific about were the samples were analysed for the three different tests. These were:

- *Chlamydia* Rapid Test: "Clinic staff tested vaginal swabs on site; all staff had passed testers' requirements in accordance with the National Committee on Clinical Laboratory Standards."
- Polymerase chain reaction assay: "We sent urine specimens to a laboratory accredited by the UK Accreditation Service for testing for Chlamydia trachomatis with the Amplicor Chlamydia trachomatis polymerase chain reaction assay."
- Transcription mediated assay: "Samples that yielded discordant results between the Chlamydia Rapid Test and the polymerase chain reaction assay were tested by transcription mediated assay at the Sexually Transmitted Bacteria Reference Laboratory."

Is the study valid? Testing

Was the reference standard applied regardless of the index test result?

- Yes, as already discussed, all samples were tested with both the *Chlamydia* Rapid test and polymerase chain reaction assay. With discordant samples further tested with transcription mediated assay

Was the diagnostic test validated in a second independent group of patients?

- Yes, as the test was given in three different sites, a total of three populations were tested

Is the study valid? Methods

Were the methods of the diagnostic test described in sufficient detail? Consider if descriptions of the following are included:

- Rationale for the ref standard.
- Technical specifications or references for running the index test and reference standard (e.g., including enough information that the tests could be replicated).
- Methods for calculating or comparing measures of diagnostic accuracy and statistical uncertainty (95% CI).

Is the study valid? Methods

Were the methods of the diagnostic test described in sufficient detail? Consider if descriptions of the following are included:

Rationale for the ref standard. "We assessed the performance of the Chlamydia Rapid Test in order to meet the requirements for Conformité Européenne licensure, which stipulate that the comparator test should be a "state of the art" assay and use specimens approved for the test. Participants from site 1 did not provide endocervical swabs, preventing the pooling of data from all three sites. Given this condition, we chose polymerase chain reaction testing, which is licensed for both urine and endocervical specimens, as the "gold standard" for the study. Studies of Chlamydia trachomatis polymerase chain reaction testing have shown equal performance with cervical and urine specimens, across all volumes of urine tested (<20-90 ml),16 and good reproducibility. For the genitourinary medicine clinics, endocervical specimens were additionally collected by the clinician and were tested by strand displacement amplification assay at the hospital laboratory."

Technical specifications or references for running the index test and reference standard (e.g., including enough information that the tests could be replicated) Yes, the study outlined in detail how each different type of sample was analysed for each test. See pages 2 and 3 for descriptions of sample collection, storage, and testing.

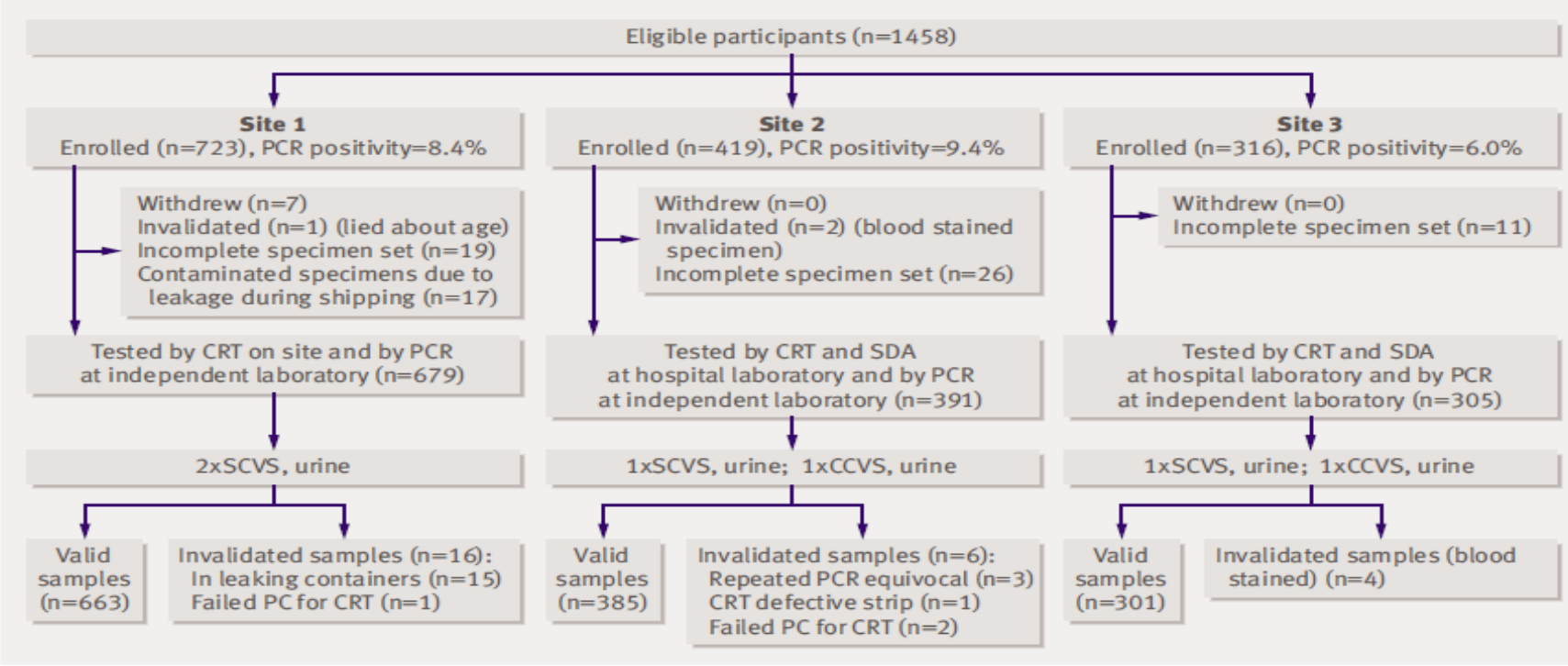
Methods for calculating or comparing measures of diagnostic accuracy and statistical uncertainty (95% CI). Yes, 95% confidence intervals were included for all comparisons discussed.

- ***Now that we have established that the study is valid, we should consider the results***

Results

Do the results include information about people who satisfied inclusion criteria for the study but did not receive the diagnostic index or reference standard test?

In this case yes, the study includes a flow chart for all three sites, which specifies how many women were enrolled and explicit reasons for any withdrawals. From this flow chart it appears that all withdrawals were excluded from the final analysis which only included valid specimens



Results

Do the results include how indeterminate results, missing results, and outliers of the index test were handled?

- The study states that samples that had discordant results were further tested by transcription mediated essay, in addition 100 of the total number of polymerase chain reaction negative specimens and 20 of the concordant positive samples were also randomly tested by the assay to minimise potential bias introduced by testing discordant samples only. The study only included valid samples in the analysis with explicit reasons for any samples not included (please see table on previous slide)

Results

Do the results include criteria for defining the severity of the target disorder?

In this case no — infection and sequelae may be asymptomatic in cases of *Chlamydia*

Results

Do the results include cross tabulation of the index test results by the reference standard results? Or enough information to generate this table?

- Yes, the study includes sensitivity, specificity, and positive and negative predictive values for all of the comparisons made, and the calculations used. Using these results, you could if needed generate the cross tabulation table, for example below:
- Site 1, *Chlamydia* Rapid Test with self collected vaginal swab specimens versus polymerase chain reaction

	Reference Standard			
Index test		Positive	Negative	Total
	Positive	47	7	54
	Negative	9	600	609
	Total	56	607	663

Results

Do the results include estimates of diagnostic test accuracy and statistical uncertainty (95% CI)?

Yes the study includes 95% CI for all comparisons made. For example:

Table 1 | Unresolved performance of *Chlamydia* Rapid Test with self collected vaginal swab specimens versus polymerase chain reaction. Values are percentages (numbers) (95% confidence intervals)

Site	Sensitivity	Specificity	Positive predictive value	Negative predictive value
1 (n=663)	83.9 (47/56) (74.3 to 93.5)	98.8 (600/607) (98.0 to 99.7)	87.0 (47/54) (78.1 to 96.0)	98.5 (600/609) (97.6 to 99.5)
2 (n=385)	80.6 (29/36) (67.6 to 93.5)	98.0 (342/349) (96.5 to 99.5)	80.6 (29/36) (67.6 to 93.5)	98.0 (342/349) (96.5 to 99.5)
3 (n=301)	83.3 (15/18) (66.1 to 100)	99.6 (282/283) (99.0 to 100)	93.8 (15/16) (81.9 to 100)	98.9 (282/285) (97.8 to 100)
Total (n=1349)	82.7 (91/110) (75.7 to 89.8)	98.8 (1224/1239) (98.2 to 99.4)	85.8 (91/106) (79.2 to 92.5)	98.5 (1224/1243) (97.8 to 99.2)

No significant difference in *Chlamydia* Rapid Test performance was apparent among three sites (P=0.278, κ statistics).

Does this diagnostic test apply to your specific patient?

Is your patient similar to the people in the study in terms of clinical and demographic characteristics?

- Yes, in this case our patient is a young woman, the study population is women 16 years and over

Is the diagnostic test available, and if so, does it reflect current practice?

- To answer this question you would need to check availability, and also how current the research is at the time of assessment

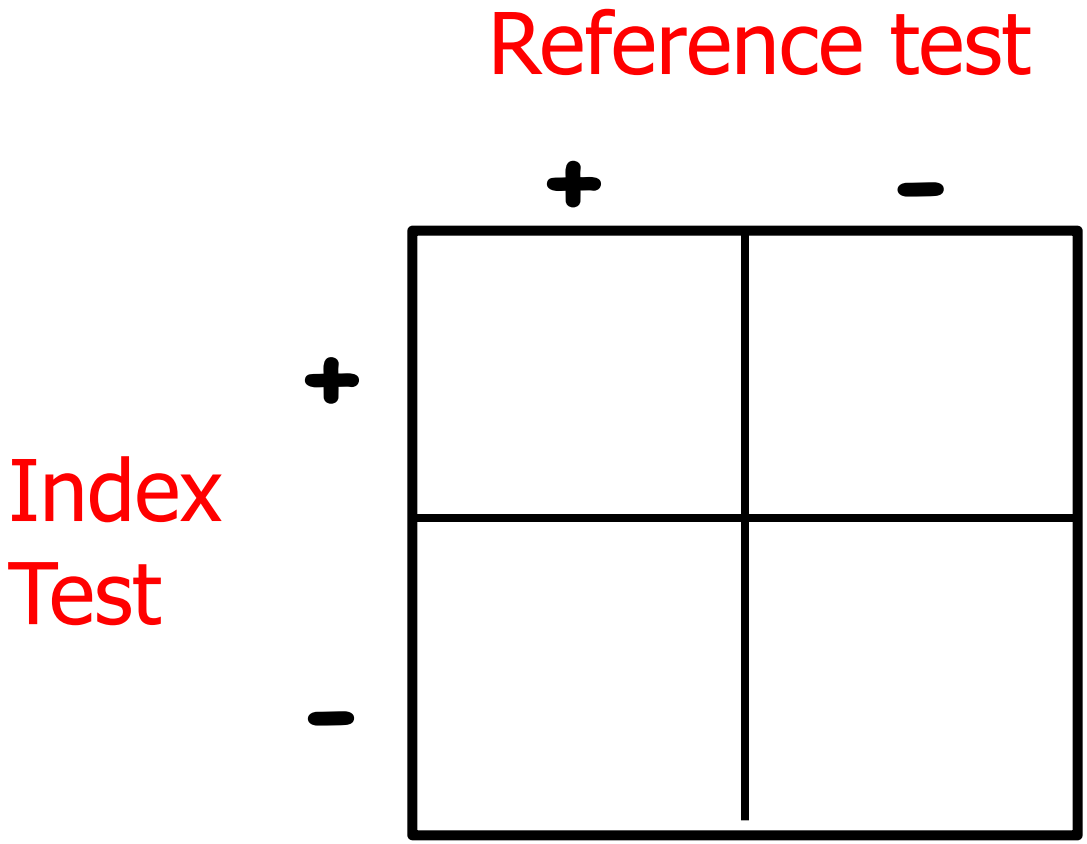
Will the test result change the way the patient is managed?

- Yes, with the Rapid *Chlamydia* test, diagnosis and treatment (if needed) is much quicker

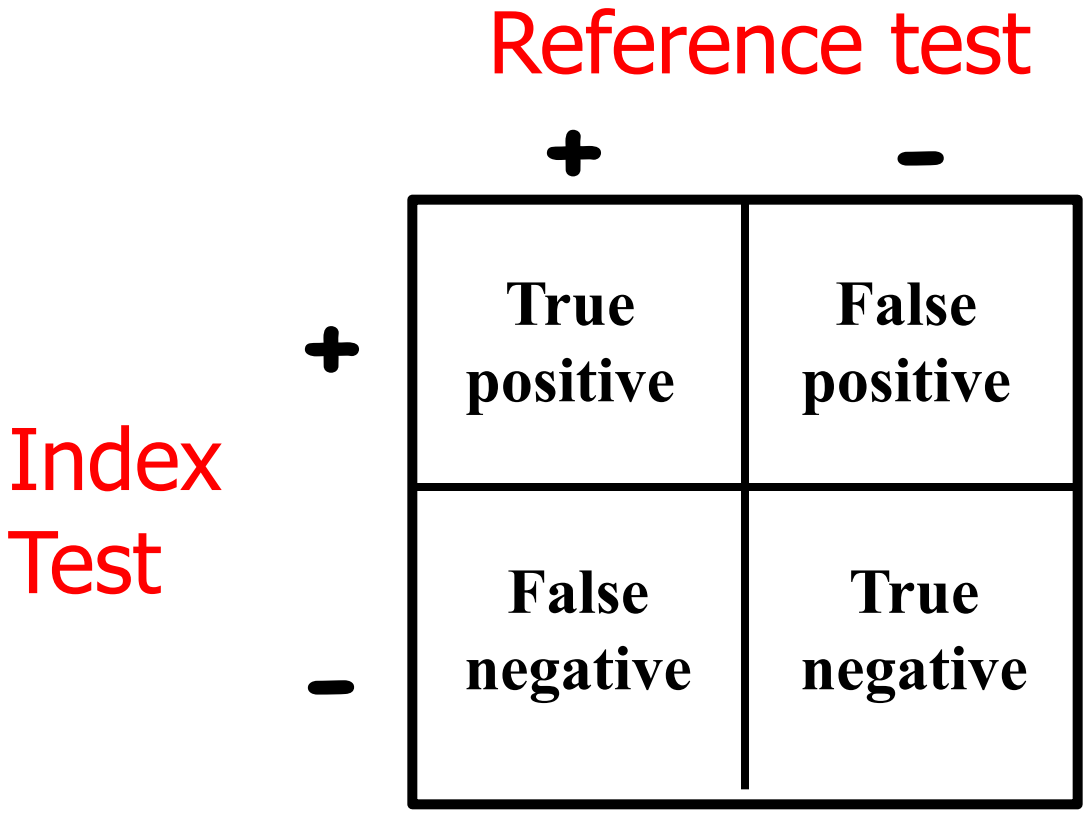
In conclusion

- This study seems to be **valid** with no major methodological flaws
- The results of the study indicate that compared with the polymerase chain reaction testing, the *Chlamydia* Rapid Test has moderate sensitivity and good specificity for screening and diagnosis of *Chlamydia* whether the vaginal swab was collected by a participant of the study or a clinician
- The study population does in this case match our patient, so we can be reasonably comfortable in the knowledge that if the patient is allowed to collect her own vaginal swab, the test result will be accurate, and also as an added bonus if the test is positive for *Chlamydia*, treatment can be started immediately

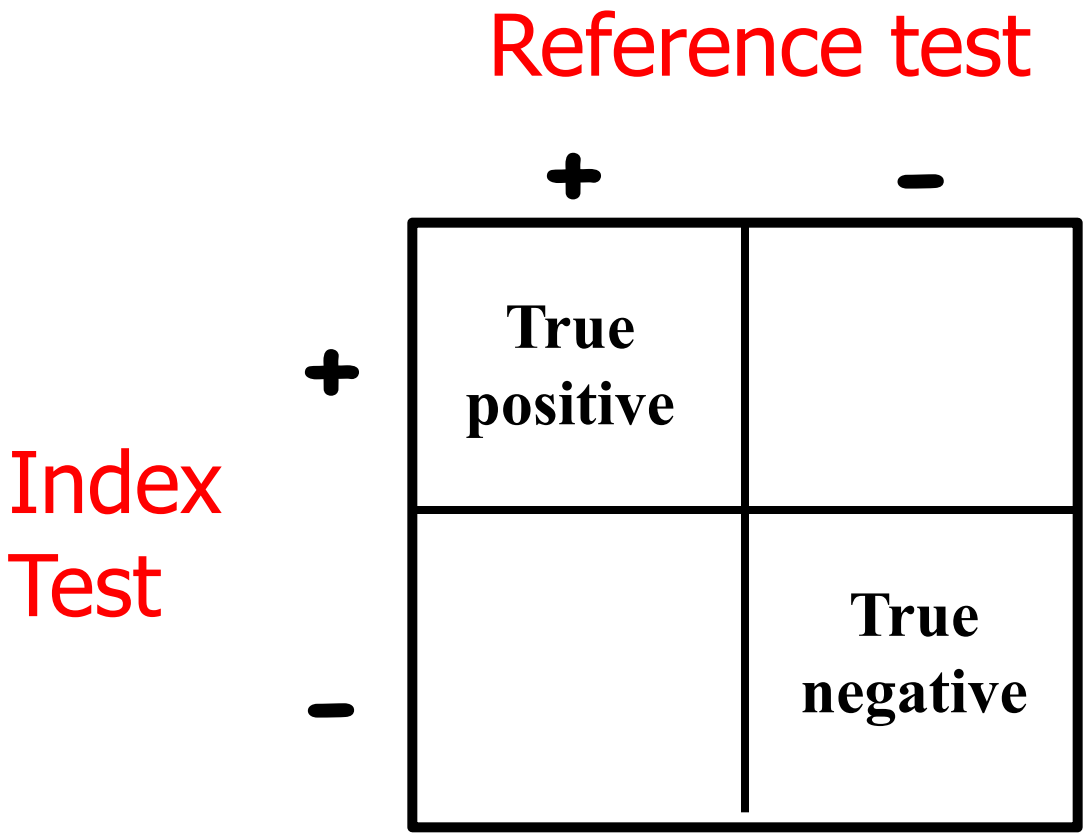
2 by 2 table



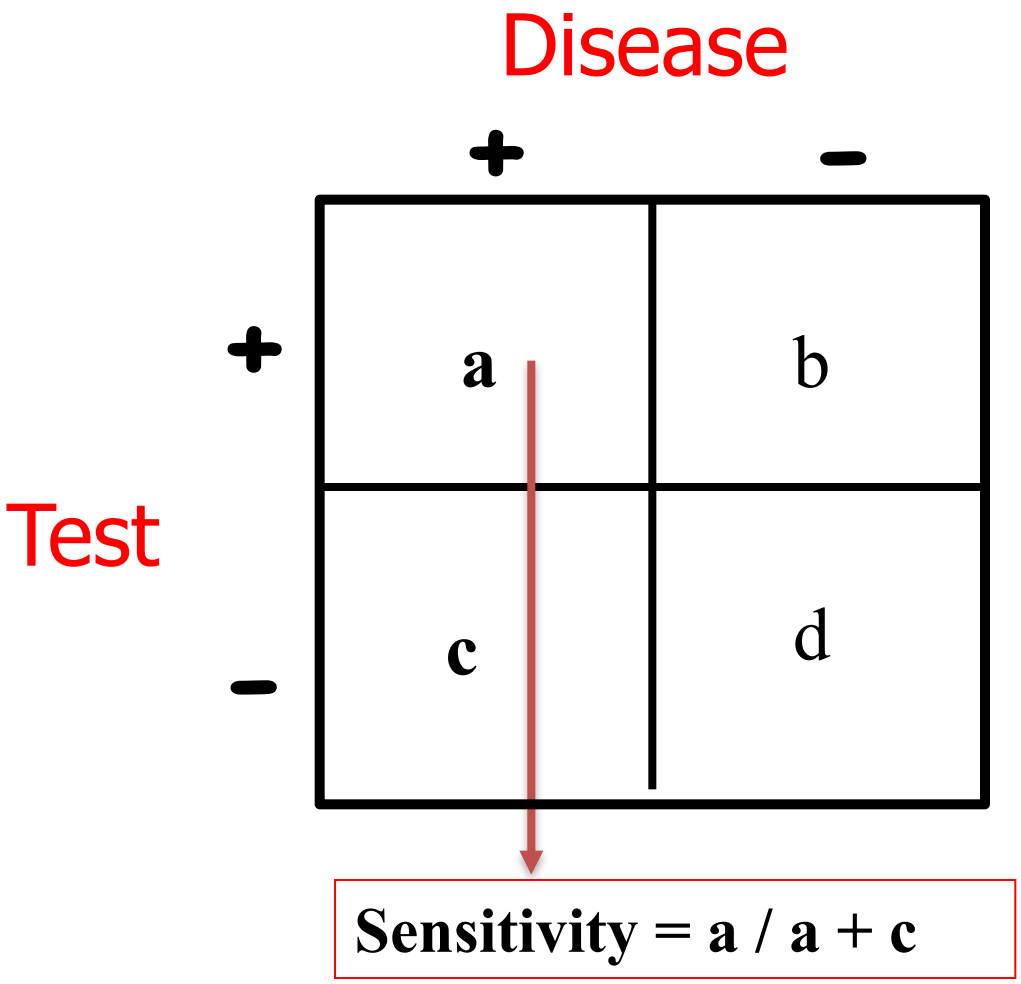
2 by 2 table


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

If only a test had perfect discrimination…

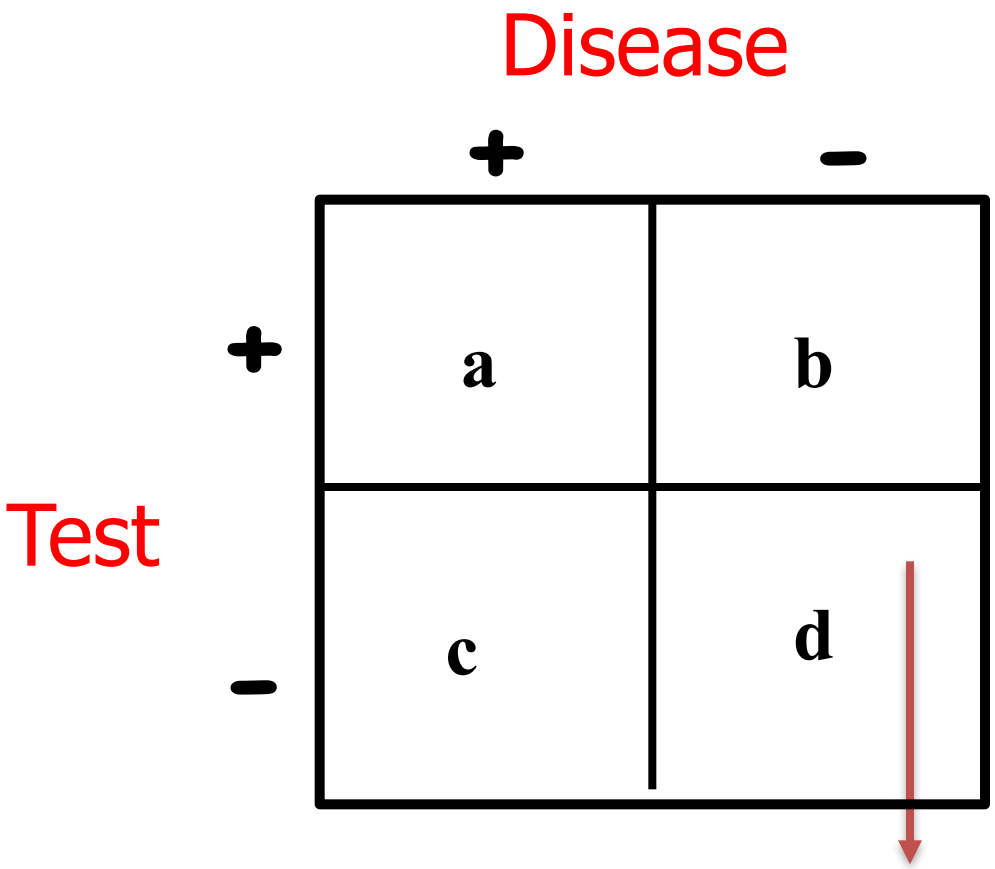


Sensitivity



Proportion of people with the disease who have a positive test.

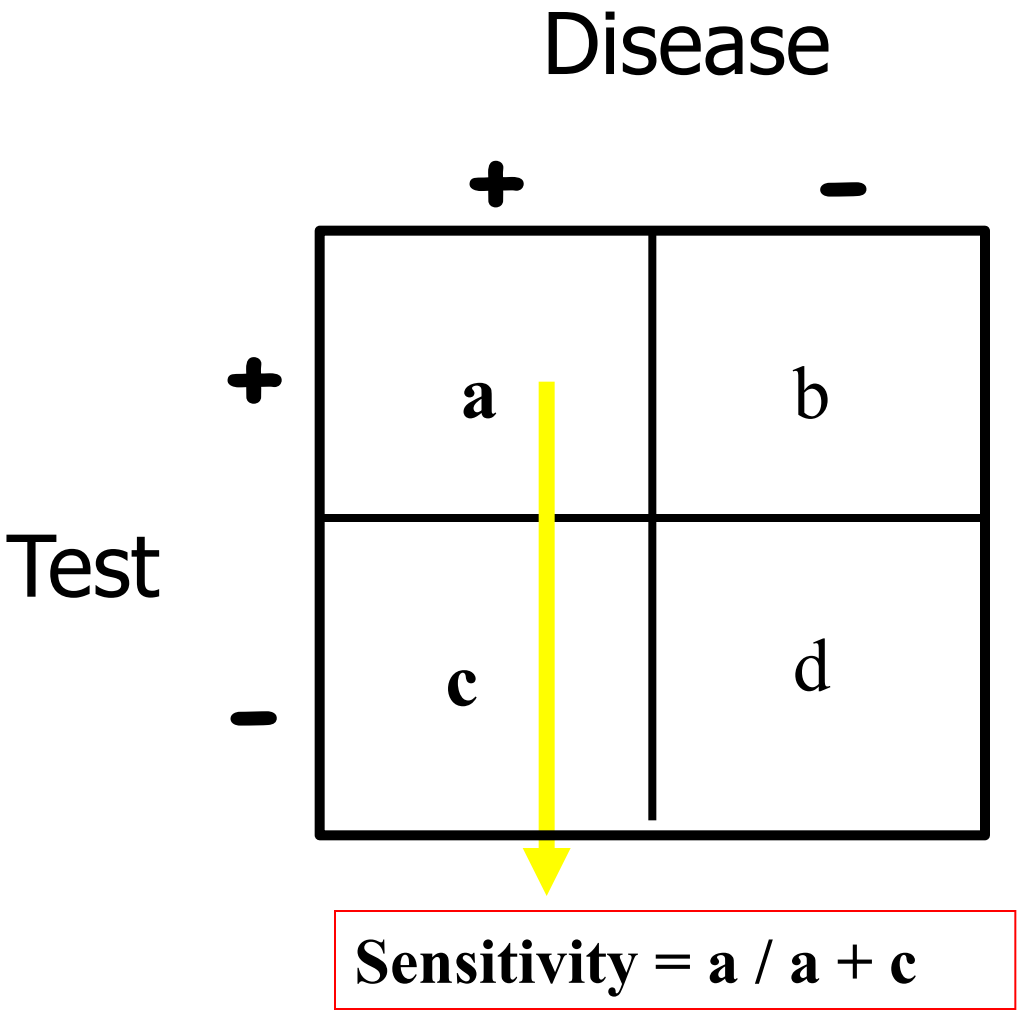
Specificity



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

Specificity = d / b + d

SnNOUT



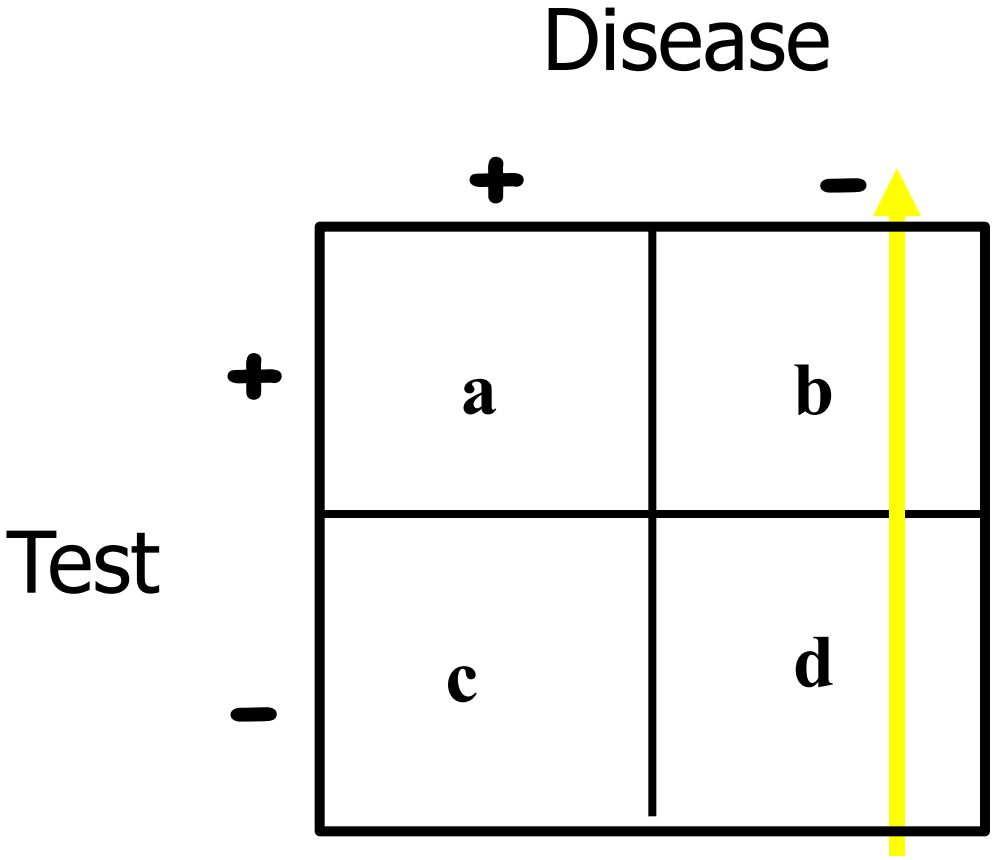
Highly sensitive tests = good for screening

or

SnNOUT

Highly sensitive test, negative result rules out.

SpPIN



Highly <u>specific</u> tests = good for ruling in

or

SpPIN

Highly specific test, positive result rules in.

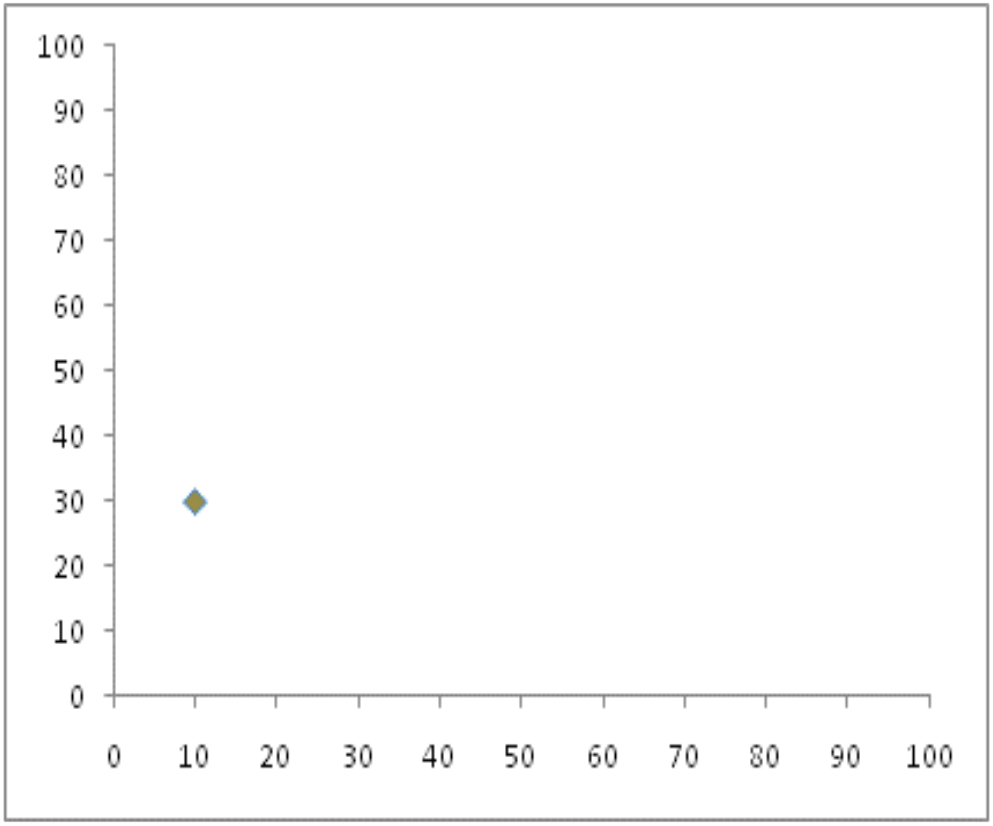
Specificity = d / b + d

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

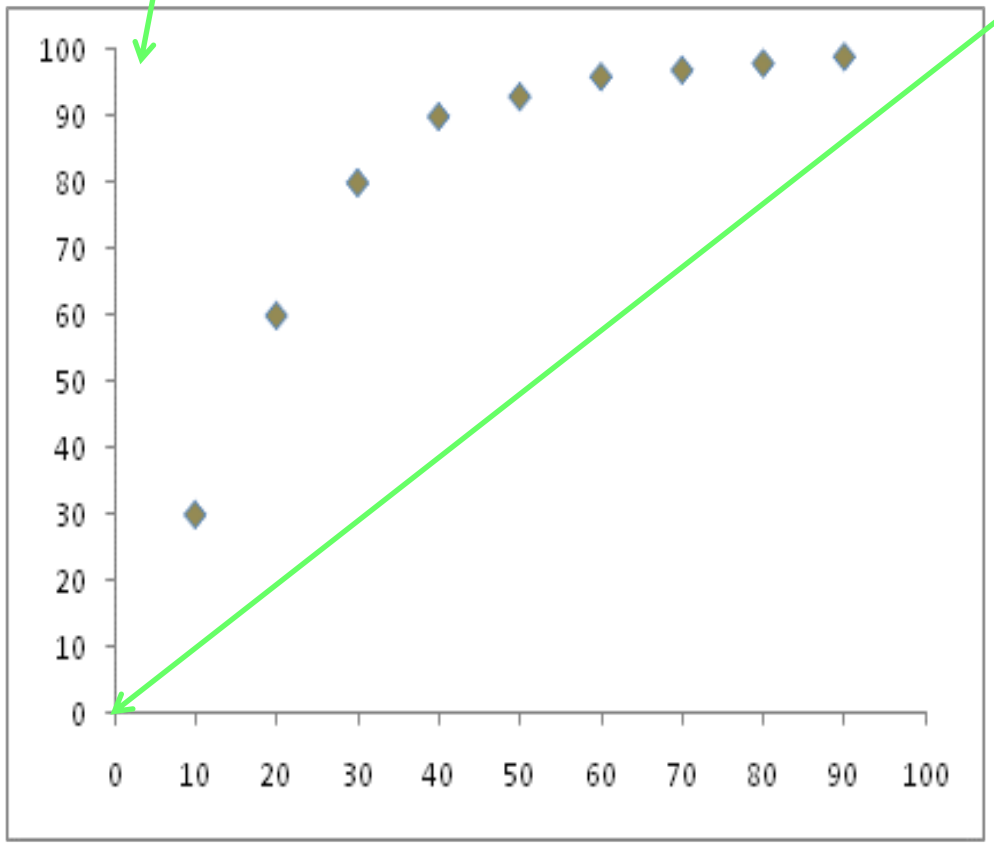


1-Specificity or false positive rate

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

Appraisal

Critical appraisal questionnaires

www.cebma.org/ebp-tools

Appraisal of a Cohort

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

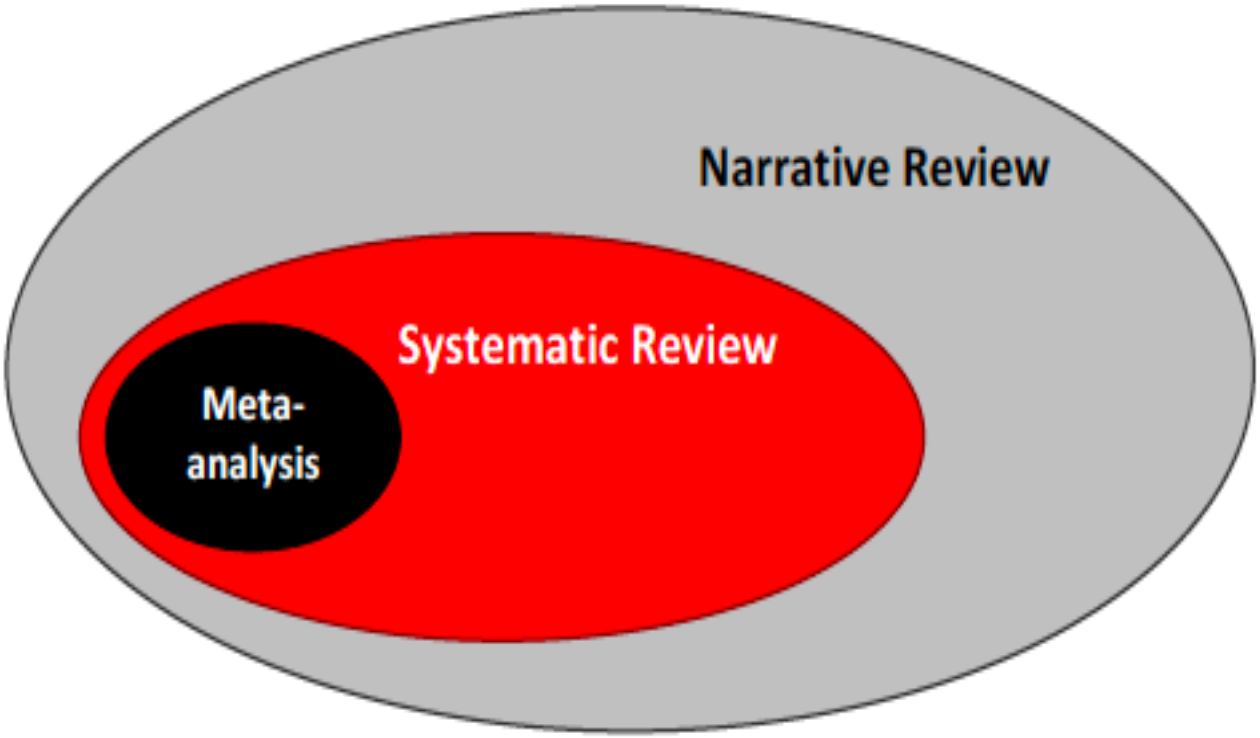
Take-home messages:

- 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?

Basics of Critical Appraisal of Available Evidence - II

(MA, SR and Guidlines)

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

Systematic review

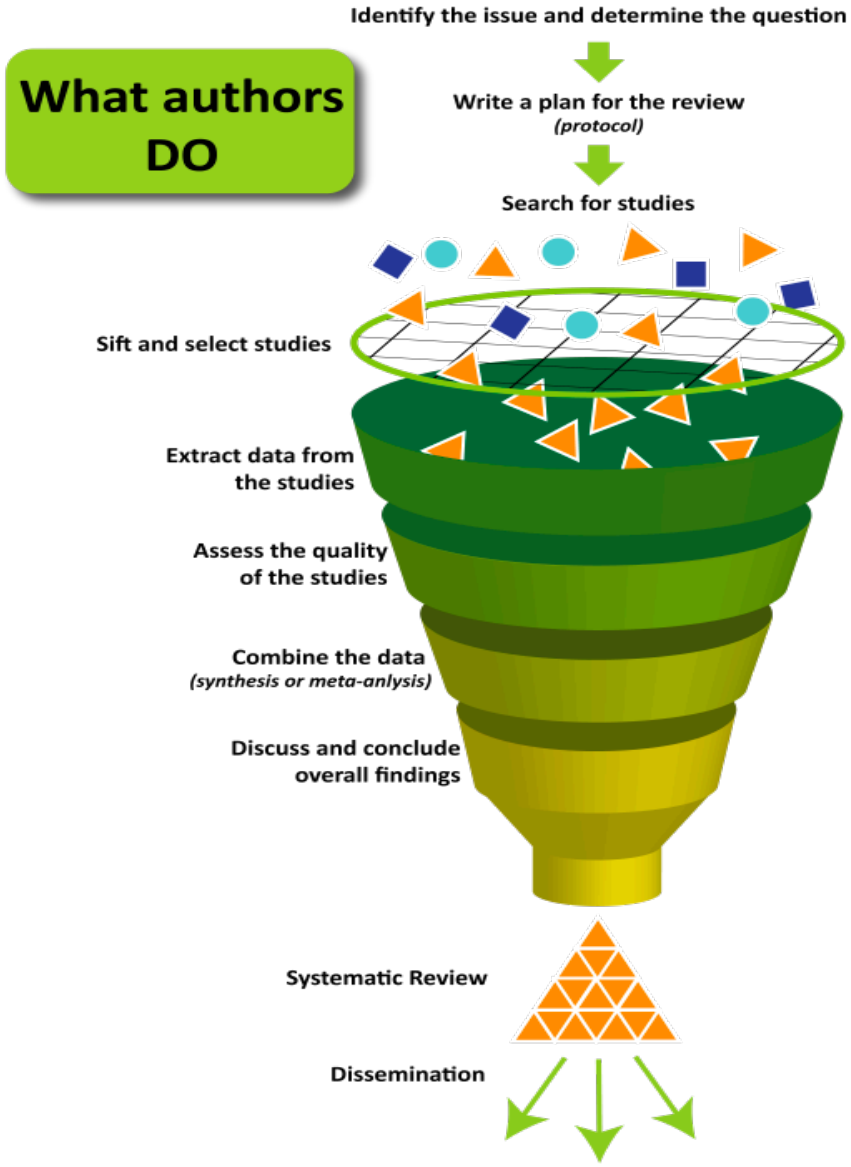
- A review of the evidence on a clearly formulated question
- Uses systematic and explicit methods to identify, select and critically appraise relevant primary research included in the review

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


What authors DO
Identify the issue and determine the question
Write a plan for the review
(protocol)
Search for studies
Sift and select studies
Extract data from the studies
Assess the quality of the studies
Combine the data
(synthesis or meta-anlysis)
Discuss and conclude overall findings
Systematic Review
Dissemination

Systematic reviews

- Scientific approach to a review article
- Criteria determined at outset
- Comprehensive search for relevant articles
- Explicit methods of appraisal and synthesis
- Meta-analysis may be used to combine data

Narrative reviews

- Depend on authors’
- inclination (bias)
- • Author gets to pick any
- criteria
- • Search any databases
- • Methods not usually
- specified
- • Vote count or narrative
- summary
- • Can’t replicate review

Advantages of systematic reviews

- Reduce bias
- Replicability
- Resolve controversy between conflicting studies
- Identify gaps in current research
- Provide reliable basis for decision making

Limitations of systematic reviews

- Results may still be inconclusive
- There may be no trials/evidence
- The trials may be of poor quality
- Practice does not change just because you have the evidence of effect/effectiveness

The Cochrane Collaboration

International nonprofit organization that prepares, maintains, and disseminates systematic up-to-date reviews 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
© Springer-Verlag 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.

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

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

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 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 (CASP**)**

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

Step 1 – Are the results of the review valid?

- Question – what is the PICO (etc.)
- Finding – comprehensive?
- Appraise – did they select good ones?
- Synthesise – 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

Tip = If cannot ascertain what the focused question is after reading these sections, search for another paper.

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

PICO:

P = patients with ACL injury = yes (initial injury suggests not chronic condition but no info on age)

I = Early ACL reconstruction = yes

C = Delayed reconstruction = yes

O = clinical and radiological = unclear

So overall we happy to say yes

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

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?

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-

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?

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

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

Were the results similar from study to study?

Consider whether

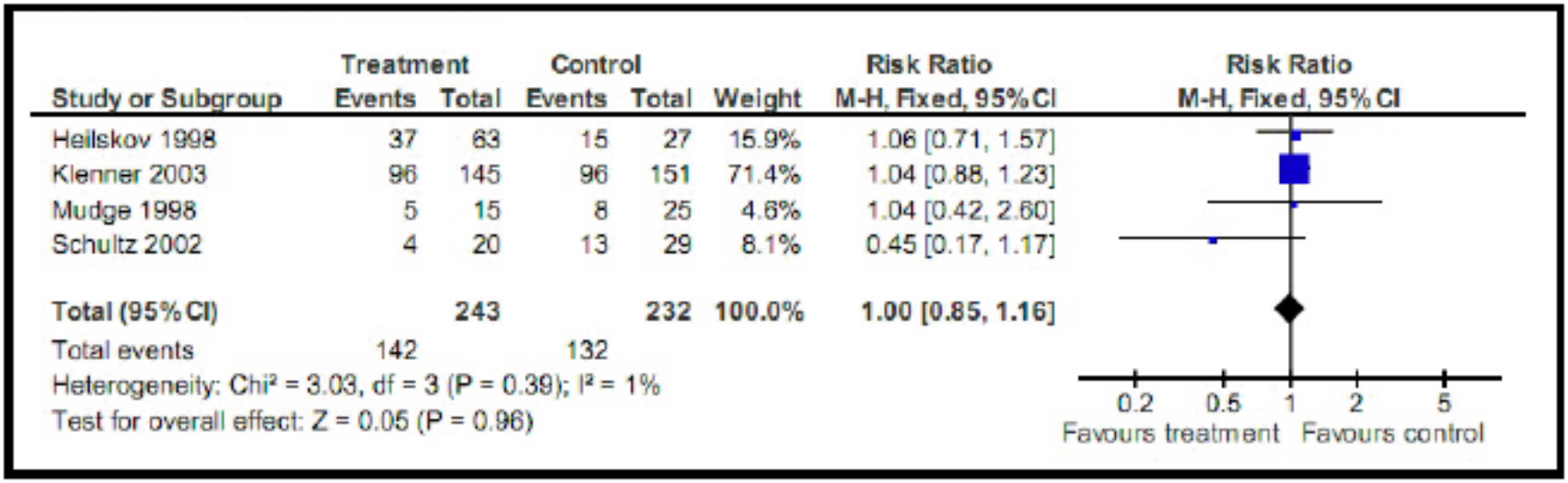
- The results of all the included studies are clearly displayed
- The results are combined (meta-analysis)
  - Are studies sufficiently similar
- The reasons for any variations in results are discussed

Meta-analysis:

- It is the pooled result and statistical methods of Systematic Review for the purpose of integrating the findings (quantitative).

Forest plot

Forest plot



Study or Subgroup	Treatment Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Heilskov 1998	37	63	15	27	15.9%	1.06 [0.71, 1.57]
Klenner 2003	96	145	96	151	71.4%	1.04 [0.88, 1.23]
Mudge 1998	5	15	8	25	4.6%	1.04 [0.42, 2.60]
Schultz 2002	4	20	13	29	8.1%	0.45 [0.17, 1.17]
Total (95% CI)		**243**		**232**	**100.0%**	**1.00 [0.85, 1.16]**
Total events	142		132			

Heterogeneity: $Chi^2 = 3.03$, $df = 3$ ($P = 0.39$); $I^2 = 1\%$

Test for overall effect: $Z = 0.05$ ($P = 0.96$)

Heterogeneity

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

Cause of heterogeneity:

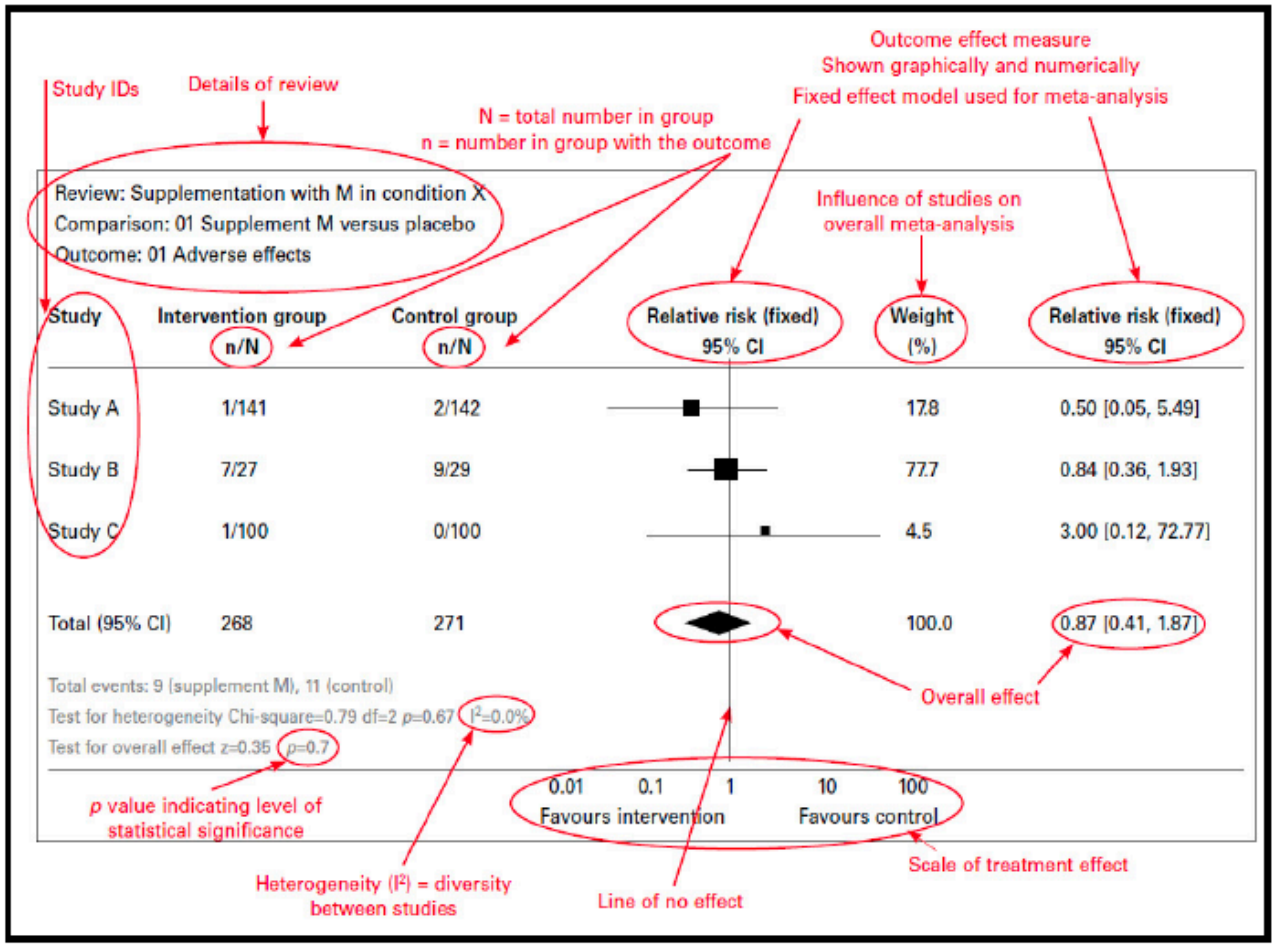
1) Clinical causes:

- Patients (age, gender, diagnosis, disease severity, inclusion and exclusion criteria).
- Intervention (type, dose, duration).
- Outcome (type, scale, cut of point, duration, follow up).

2) Methodological (sample size error, study quality, design, analysis).

3) Statistical (by chance "randomization").

Description of forest plot



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'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.

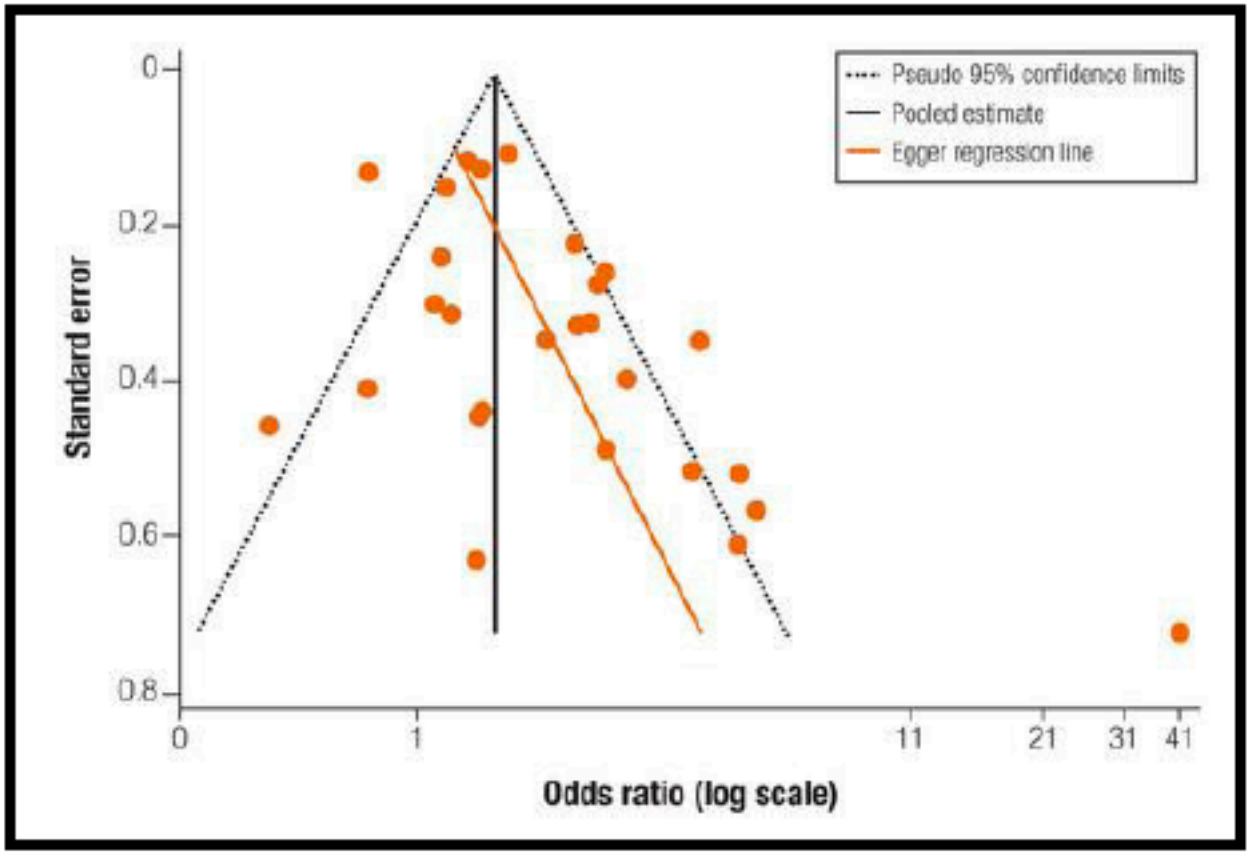
- Can I apply these results to my case?
- Is my patient so different to those in the study that the results cannot apply?

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).

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

Funnel plot


Pseudo 95% confidence limits
Pooled estimate
Egger regression line
Standard error
0
0.2
0.4
0.6
0.8
0
1
11
21
31
41
Odds ratio (log scale)

Funnel Plot

- It is visual aid to assess for publication bias if there is asymmetrical distribution around pooled result line.
- Y-axis line represents the standard error of the mean. The lesser standard error, the larger the sample size; i.e. the upper most studies in funnel plot are the bigger in sample size.
- X-axis line represents the estimated effect size (OR or RR).
- The vertical line in the middle of the funnel plot represents the "pooled estimate" result.
- The distance from the "pooled estimate" line and the border of funnel (dashed lines) represent the 95% CI; the more distance the study from the "pooled estimate" line the wider CI of that study.

Funnel Plot

- 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


Thank
you!!

Refrences

- 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/

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/