

Critical Review of Medical Literature

Irina Burd, MD, PhD
Maternal Fetal Medicine
Johns Hopkins University

Is the study valid?

- Research question
- Population
- Intervention
- Analysis

What is the research question?

- Is there a well stated hypothesis?
- Does it include a clinically meaningful and relevant outcome?
- What is the end point?
 - Real
 - Surrogate end point=biomarker that is intended to substitute for a clinical end point
 - Usually used when outcome is rare, long time to observe the outcome, difficult to measure outcome, rapid dissemination of new treatment such as Phase II and III trials

Is the study valid?

Research question

Population

Intervention

Analysis

What is the study sample?

- Population → sample
- FLOW CHART/Figure 1 available and clear
- Table 1 is available to describe the sample
- Is sample representative of source population?
- Is sample/population representative of population in general
- Is inclusion and exclusion criteria well stated?
- ULTIMATELY, “CAN I USE THIS STUDY AND APPLY TO MY POPULATION?”

Is the study valid?

Research question

Population

Intervention

Analysis

Intervention

What kind of study is this?

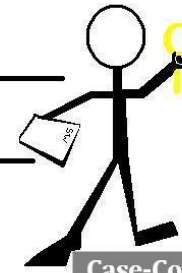
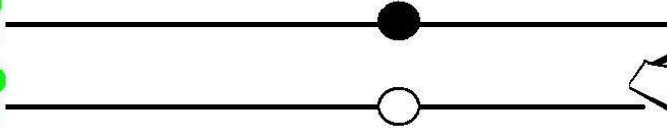
- Case series
- Case control
- Cohort
- Randomized control study

Observational Study Designs: Case Control vs Cohort

Wu, S.
Brainfairs

Exposure

Disease

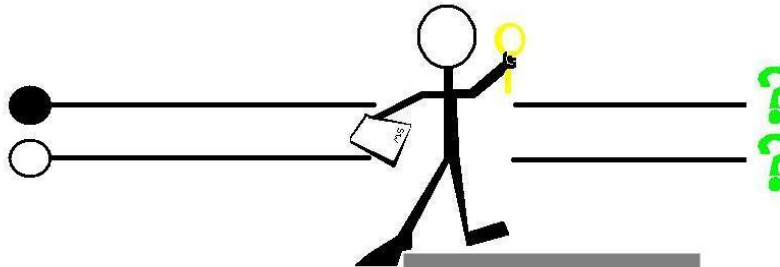


Case-Control

Can't use RR, can only use OR because researcher sets the prevalence within the study. Good for rare diseases. In rare diseases, OR approximates RR. In non-rare diseases, the direction of OR and RR are the same, but the actual number obtained for OR and RR are different. You CANNOT obtain a RR for this. It makes no sense to.

Exposure

Disease



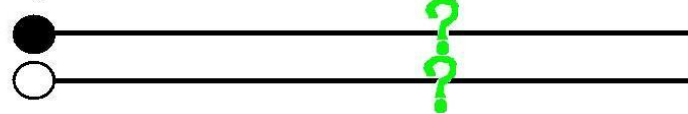
Prospective Cohort

RR and OR are both relevant for this. This is sometimes used to test out a new intervention/treatment.

RR and OR are both relevant for retrospective cohorts.

Exposure

Disease



Retrospective Cohort



Investigator/Researcher begins their research. When the researcher enters the scene.

KEY



Present



Absent



What we are seeking; the information we are trying to obtain; what we do not know; our question.

Is the study valid?

Research question

Population

Intervention

Analysis

Analysis

- Is there a right analysis for the data available
 - Power
 - Multiple comparisons

Is power analysis for the comparison of groups necessary and is it done?

- Power analysis
 - Beta, the Type II error rate, must be kept low
 - Power which is equal to $1 - \beta$, must be kept correspondingly high
 - Ideally, power should be at least 0.80 to detect a reasonable departure from the null hypothesis.

Results

Univariate analysis	Frequency tables (usually, Table 1)
Bivariate	2 by 2 tables (Sensitivity, Specificity, PPV, NPV)
Time to event	Kaplan Meier analysis
Multivariate	Depends on type of data: Binary – Logistic regression Continuous - Linear regression/analysis of variance Time to event: Cox proportional hazards models

Calculation of effect size

- Relative risk
 - Estimates the magnitude of association between exposure and disease in exposed relative to unexposed
- Odds ratio
 - Ratio of odds of exposure among cases to odds of exposure among controls
- When to use one versus the other?

Relative risk (RR)

	Disease	No disease
Exposure	a	b
No exposure	c	d

$$RR = \frac{\text{Risk of disease in exposed}}{\text{Risk of incidence of disease in unexposed}} = \frac{a/(a+b)}{c/(c+d)}$$

RR=1 There is no association between exposure and disease state

RR>1 There is an increased risk of disease in exposed

RR<1 There is a decreased risk of disease in exposed

Odds ratio

	Disease	No disease
Risk factor present	a	b
Risk factor absent	c	d

$$OR = \frac{ad}{bc}$$

OR=1 The events are equally likely in 2 groups

OR>1 Event is more likely in one group

OR<1 Event is less likely in one group

How large is the treatment effect

- Risk without therapy
 - Baseline risk **X**
- Risk with therapy **Y**
- Absolute risk reduction = **X-Y**
- Relative risk = **Y/X**
- Relative risk reduction = **[(X-Y)/X]/100**

BIAS

- What is bias?
- Are different biases considered?
 - Selection bias, including:
 - Sampling bias is systematic error due to a non-random sample of a population, causing some members of the population to be less likely to be included than others, resulting in a biased sample
 - Time interval/length time bias
 - Analyzing the lengths of intervals by selecting intervals that occupy randomly chosen points in time or space, a process that favors longer intervals
 - Attrition bias is a kind of selection bias caused by attrition (loss of participants), discounting trial subjects/tests that did not run to completion. It includes dropout, non-response, withdrawal and protocol deviators. Therefore, looking at the proper follow up is important
 - Lead-time bias
 - is the bias that occurs when two tests for a disease are compared, and one test (the new, experimental one) diagnoses the disease earlier, but there is no effect on the outcome of the disease-- it may appear that the test prolonged survival, when in fact it only resulted in earlier diagnosis when compared to traditional methods
 - Observer bias
 - Researcher unconsciously influences the participants of an experiment. It is a significant threat to a study's internal validity and is therefore typically controlled by using a double-blind experimental design
 - Recall bias
 - In surveys

Is it a diagnostic testing or a screening question?

Key Concept

- Sensitivity
 - The proportion of people who truly have a designated disorder who are so identified by the test.
 - **S**ensitive tests have few false **n**egatives.
 - When a test with a high **S**ensitivity is **N**egative, it effectively rules **o**ut the diagnosis of disease: **SnNout**.

Key Concept

- Specificity
 - The proportion of people who are truly free of a designated disorder who are so identified by the test.
 - **S**pecific tests have few false **p**ositives.
 - When a test is highly **S**pecific, a **P**ositive result can rule **in** the diagnosis: **SpPin**.

Key Concepts

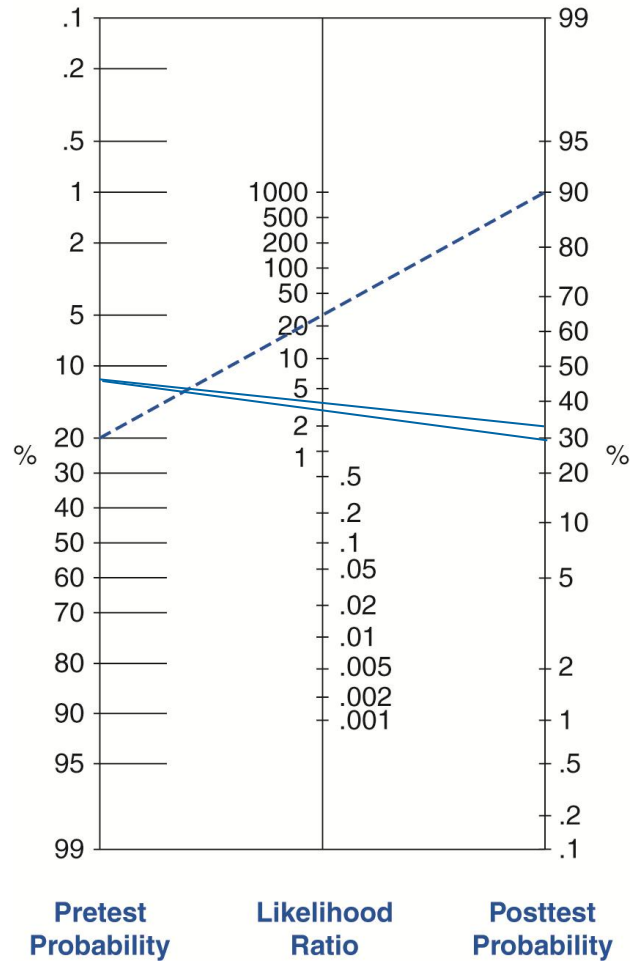
- Negative predictive value
 - The probability that the disease is not present given a negative test
- Positive predictive value
 - The probability that the disease is present given a positive test

Key Concept

- Likelihood Ratio
 - Relative likelihood that a given test would be expected in a patient with (as opposed to one without) a disorder of interest.

$$\text{LR} = \frac{\text{probability (\%)} \text{ of a } \mathbf{\text{test result}} \text{ in patients } \mathbf{\text{WITH}} \text{ disease}}{\text{probability (\%)} \text{ of the } \mathbf{\text{test result}} \text{ in patients } \mathbf{\text{WITHOUT}} \text{ disease}}$$

How Do I Use the LR?



Importance

- What likelihood ratios were associated with the range of possible test results?
 - How much will different levels of the diagnostic test raise or lower the pretest probability of disease?

What likelihood ratios were associated with the range of possible test results?

Pulmonary Embolism

Scan Results	Present		Absent		Ratio
	Number	Proportion	Likelihood Number	Proportion	
High probability	102	$102/251 = 0.406$	14	$14/630 = 0.022$	18.3
Intermediate probability	105	$105/251 = 0.418$	217	$217/630 = 0.344$	1.20
Low probability	39	$39/251 = 0.155$	273	$273/630 = 0.433$	0.36
Normal/near normal	5	$5/251 = 0.020$	126	$126/630 = 0.200$	0.10
Total	251		630		

Importance

Comparison of the Results of a Diagnostic Test With the Results of Reference Standard Using a 2×2 Table

Test Results	Reference Standard	
	Disease Present	Disease Absent
Test positive	TP	FP
Test negative	FN	TN

$$\text{Sensitivity (Sens)} = \frac{TP}{TP + FN}$$

$$\text{Specificity (Spec)} = \frac{TN}{FP + TN}$$

$$\text{Likelihood ratio for positive test (LR+)} = \frac{\text{Sens}}{1 - \text{Spec}} = \frac{\text{True positive rate}}{\text{False positive rate}} = \frac{TP / (TP + FN)}{FP / (FP + TN)}$$

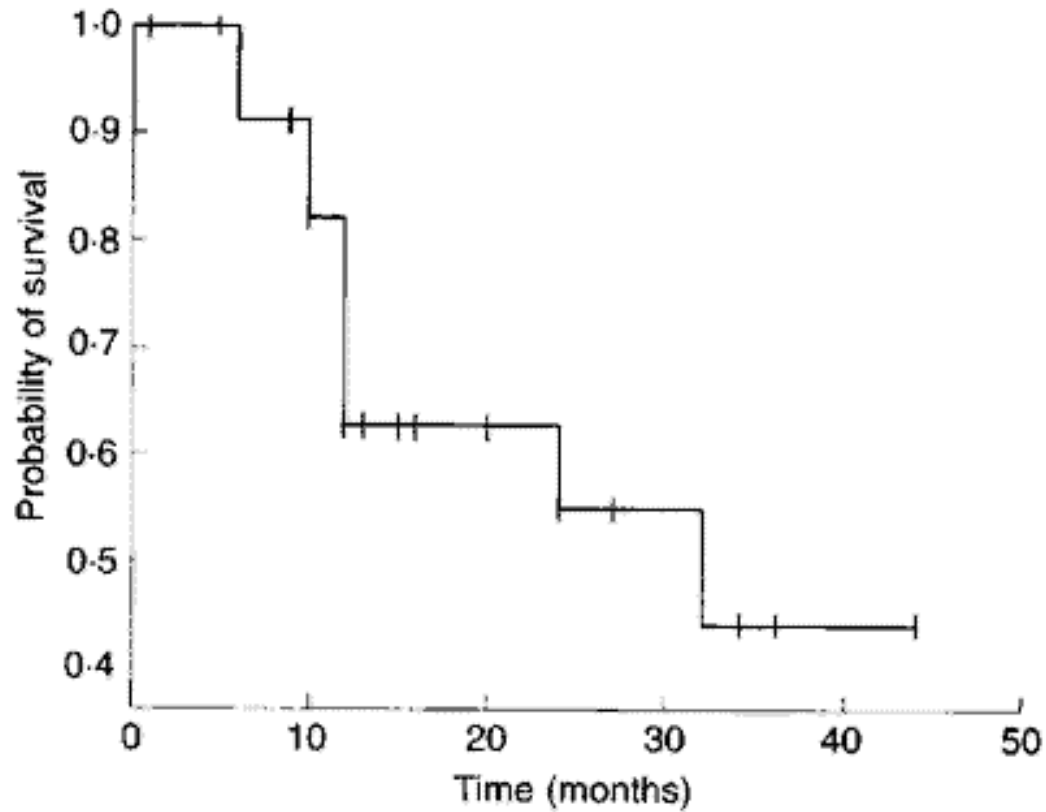
$$\text{Likelihood ratio for negative test (LR-)} = \frac{1 - \text{Sens}}{\text{Spec}} = \frac{\text{False negative rate}}{\text{True negative rate}} = \frac{FN / (TP + FN)}{TN / (FP + TN)}$$

Is it a study on prognosis of
disease?

Key concepts

- Prognosis
 - Possible outcomes of a disease and the frequency with which they can be expected to occur
- Survival curve
 - A curve that starts at 100% of the study population and shows the population still surviving (or free of disease) at successive times for as long as the information is available

Survival curve



Key concept

- Hazard ratio (time to an event rather than an event itself)
 - A hazard is the instantaneous risk of an event at a given point
 - hazard ratio can be an estimate of relative risk
- Interpreting hazard ratios
 - $HR < 1$ is protective
 - $HR > 1$ is risk factor
- In survival analysis, HR is the effect of one variable on the hazard or risk of an event

Key concept

- Cox Hazard model
 - Estimate of the treatment effect on survival after adjustment has been made for other variables
 - It allows for an estimation of the hazard or a risk of death (or another event of interest), given the patient's prognostic variables

Is it a treatment question?

Key concepts

– Relative risk reduction

- Ratio of risk in exposed subjects divided by the risk in unexposed subjects

– Number needed to treat

- The number of patients who need to be treated over a specific period of time to prevent one bad outcome

When looking at results

- Consider how precise are the estimates
 - how wide are the confidence intervals

Applicability

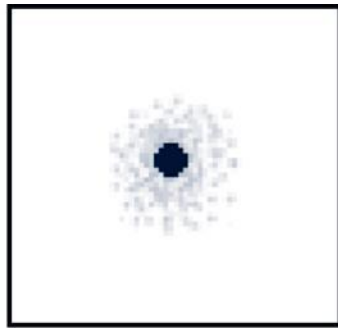
- Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?
 - Does the test yield the same result when reapplied by the same observer (intraobserver variability)?
 - Do different observers agree about the test result (interobserver variability)?

Applicability

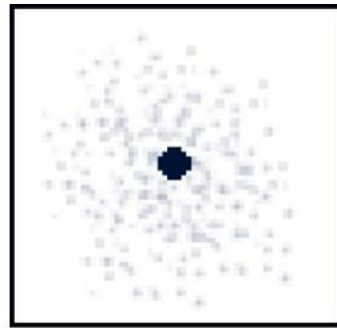
- Are the results applicable to the patients in my practice?
 - Does the test perform differently (different LRs) for different severities of disease?
 - Does the test perform differently for populations with different mixes of competing conditions?

Applicability

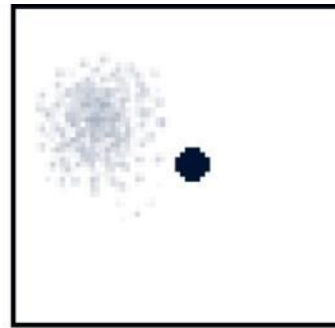
- Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?
 - Difference between validity and reliability



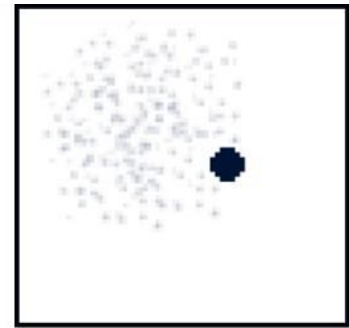
A



B



C

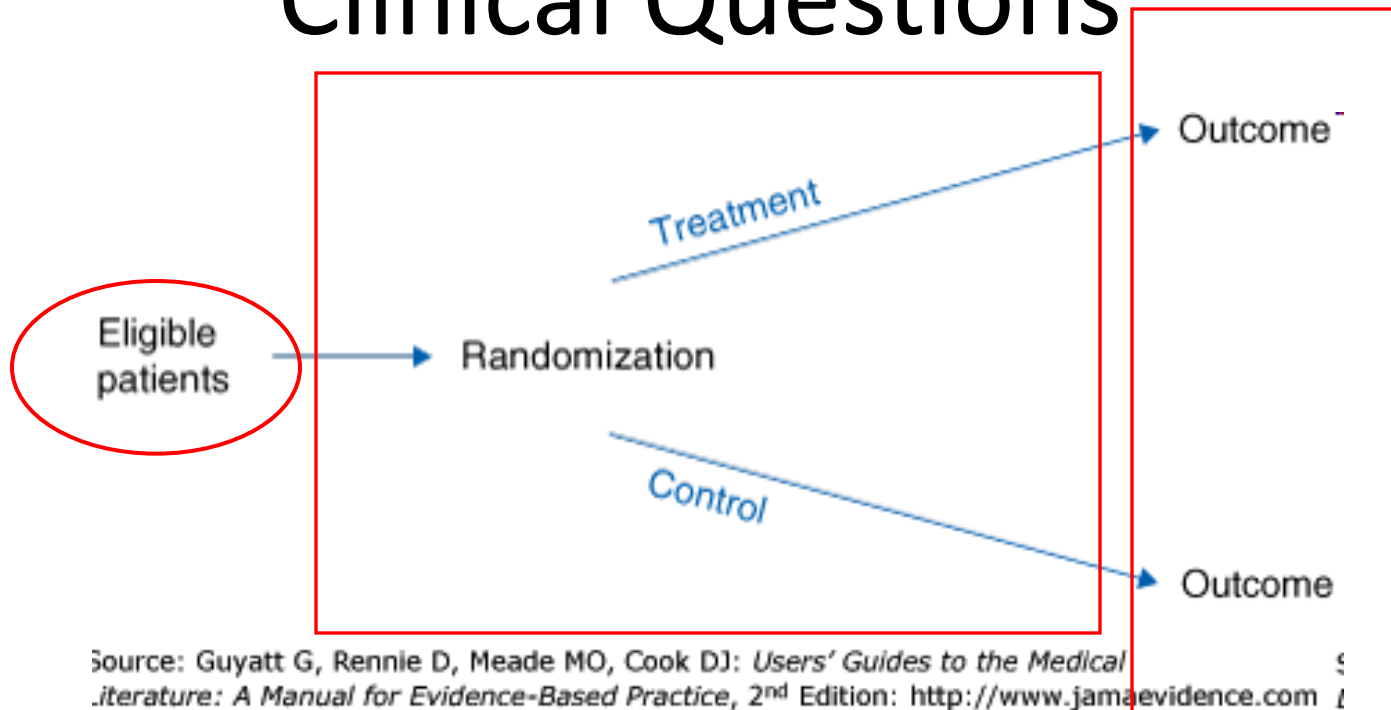


D

Themes in the Medical Literature

- How do we use the medical literature?
 - Browsing
 - Problem Solving
 - Background and Foreground Questions

Clinical Questions



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>

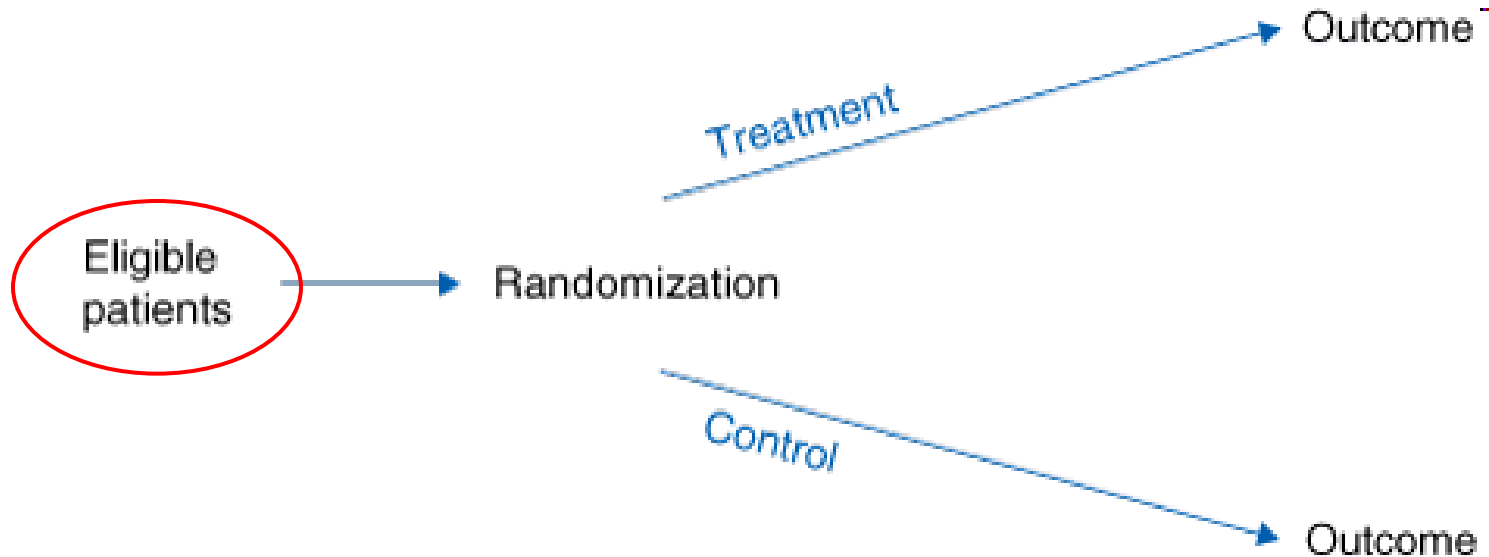
Copyright © American Medical Association. All rights reserved.

- Clinical Population -> Study Population
- Intervention/Outcomes -> Methods
- Results -> Clinical Population

Population

- What barriers exist to going from a clinical population to a study population?
- How can we ensure that the study population represents the true population of interest?

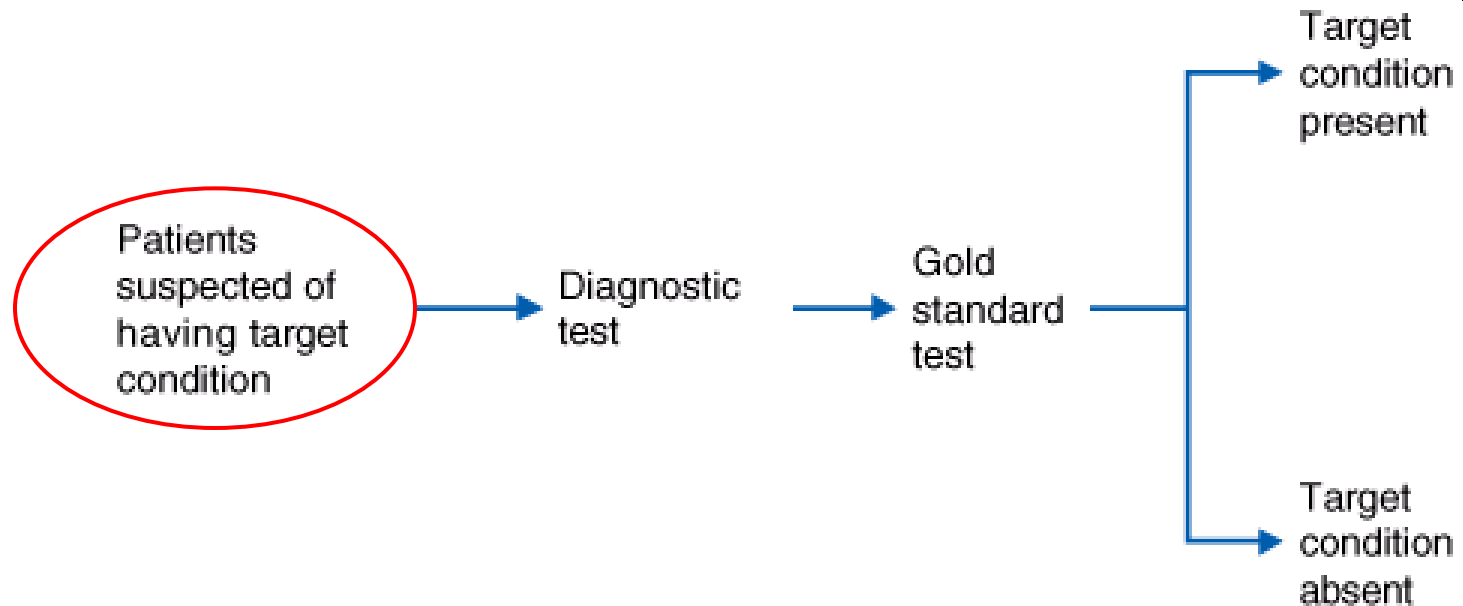
Therapy Study



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- Did intervention and control groups start with the same prognosis?
- What were the inclusion / exclusion criteria?
- Where did recruitment occur?

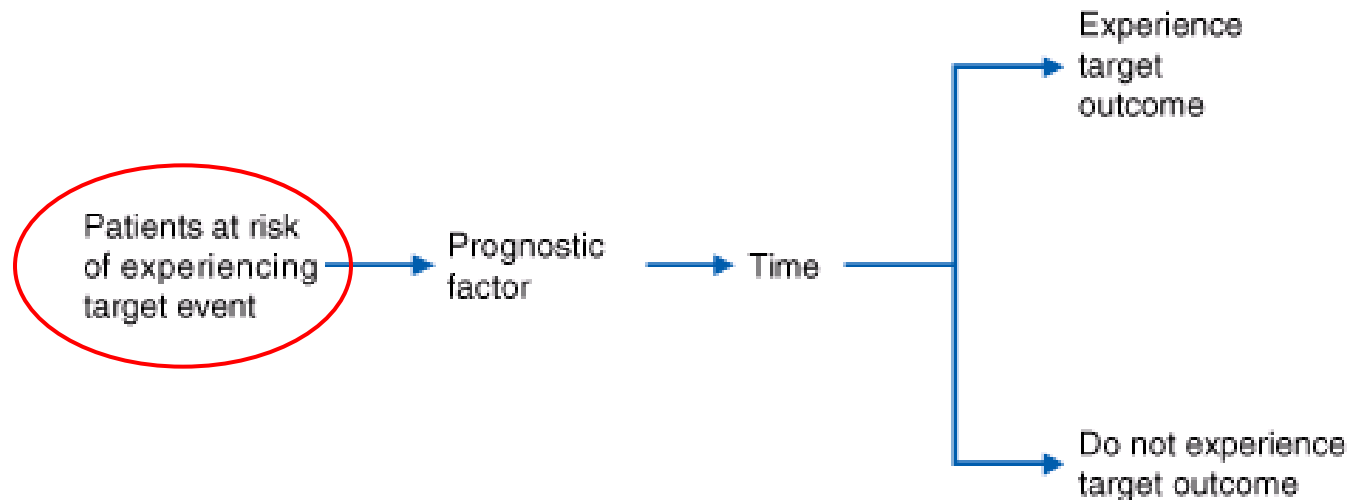
Diagnostic Test



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- Does the sample of patients represent the full spectrum of patients with and without the diagnosis of interest?
- Did participating patients present a diagnostic dilemma?
 - Were subjects drawn from a group in which it is not known whether the condition of interest is present or absent?
- Does pre-test probability fall between test and treatment thresholds?

Prognosis Study



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- Is the cohort representative of a defined group or population?
 - Referral bias, failure to clearly define study patients
 - Must report implicit filters passed before entering the study. Describe inclusion/exclusion criteria.
- Were the patients sufficiently homogenous with respect to prognostic risk?
 - Patients should be at similar point in disease process.

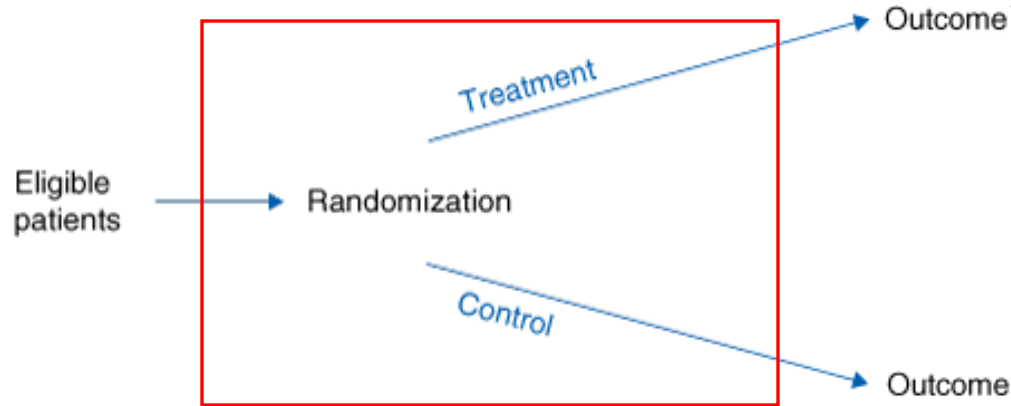
Other Considerations

- Qualitative Study
 - Were participants relevant to the research question?
 - Was participant selection well reasoned?
 - Did selection criteria evolve over the course of sampling after investigators returned to the data to explore new ideas?

Interventions/Exposures/Outcomes

- What considerations should we apply to methods and results?

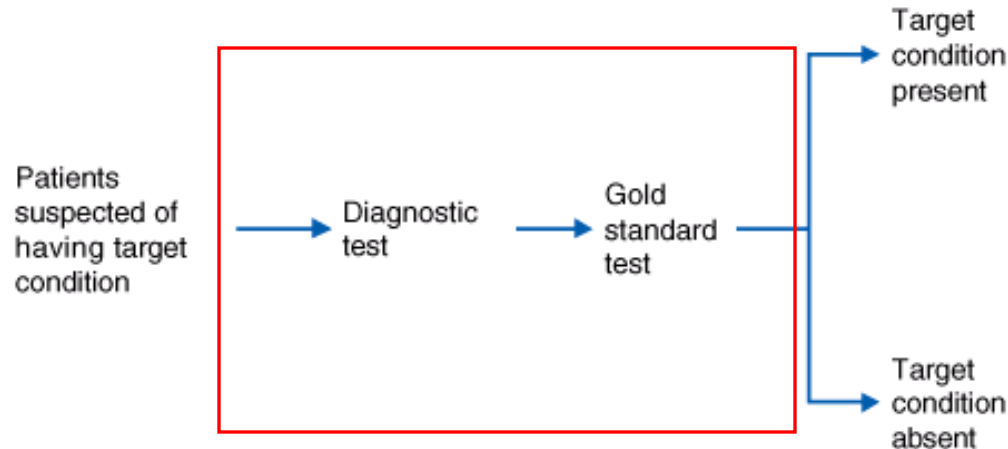
Therapy Study



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- Did intervention and control groups start with the same prognosis?
 - Were patients randomized?
 - Was randomization concealed?
 - Were patients in the treatment and control group similar with respect to known prognostic variables?
- Was prognostic balance maintained during study?
 - Were participants, researchers, and outcome assessors blinded to group allocation throughout?
- Was prognostic balance maintained at study completion?
 - Was follow-up complete?
 - ITT analysis?
- Was trial stopped early?

Diagnostic Test



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- Did investigators compare the test to an appropriate, independent reference standard?
 - Was the reference standard acceptable?
 - Were the test and standard independent?
 - Were the test and standard results assessed blindly?
- Were those interpreting the test and reference standard blind to the other results?
 - Is the test result interpretation subjective? Expectation bias?
- Did investigators perform the same reference standard to all patients regardless of the results of the test under investigation?
 - Was a different reference standard applied to subjects testing negatively?

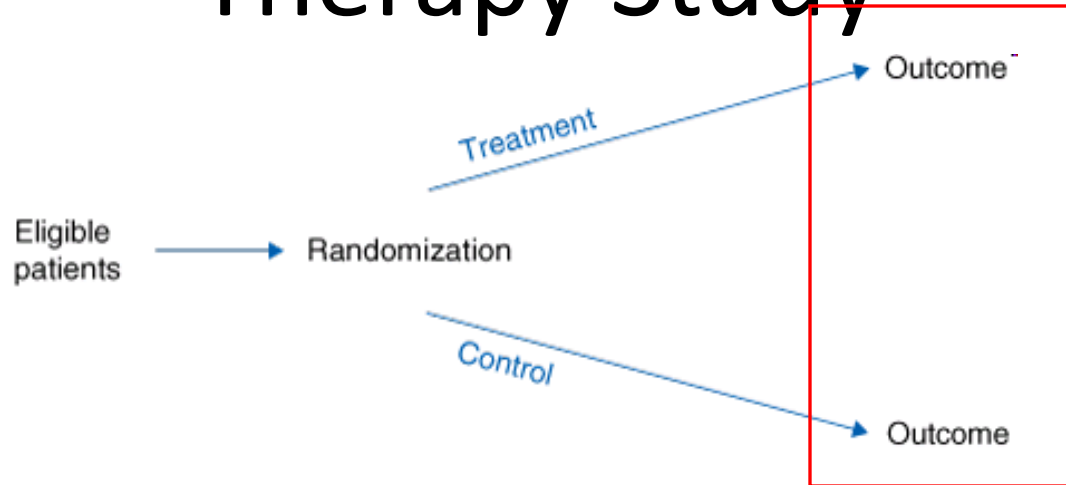
Cost-Effectiveness Study

- What viewpoint is used to evaluate costs and consequences (patient, hospital, third-party payer, or society at large)?
- What outcomes are being compared? Are costs or charges being evaluated?
- Where are estimates coming from and are they realistic?
- Did investigators adopt a sufficiently broad viewpoint?
- Are results reported separately for patients whose baseline risks differ?
- Were costs measured accurately?
- Did investigators consider the timing of costs and consequences?

Qualitative Study

- Is data collection comprehensive? Iterative data collection and analysis?
- Were the data appropriately analyzed and the findings adequately corroborated?
- Was theoretical saturation achieved and did data collection and analysis yield meaningful descriptions of social phenomena?

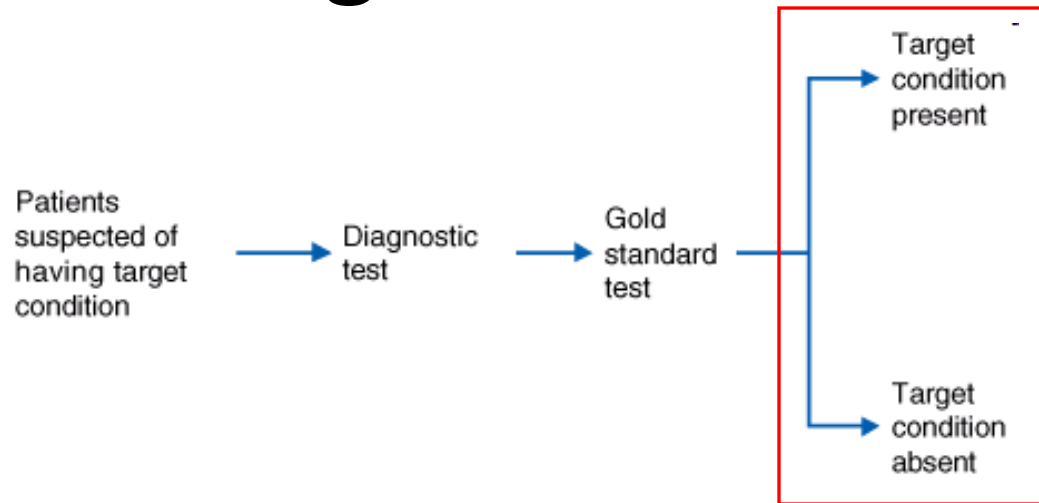
Therapy Study



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- How large was the treatment effect?
- How precise was the estimate of the treatment effect?
- How can I apply the results to patient care?
 - Study patients similar to my patient?
 - Were all patient important outcomes considered?
 - Are the likely treatment benefits worth the potential harms and costs?

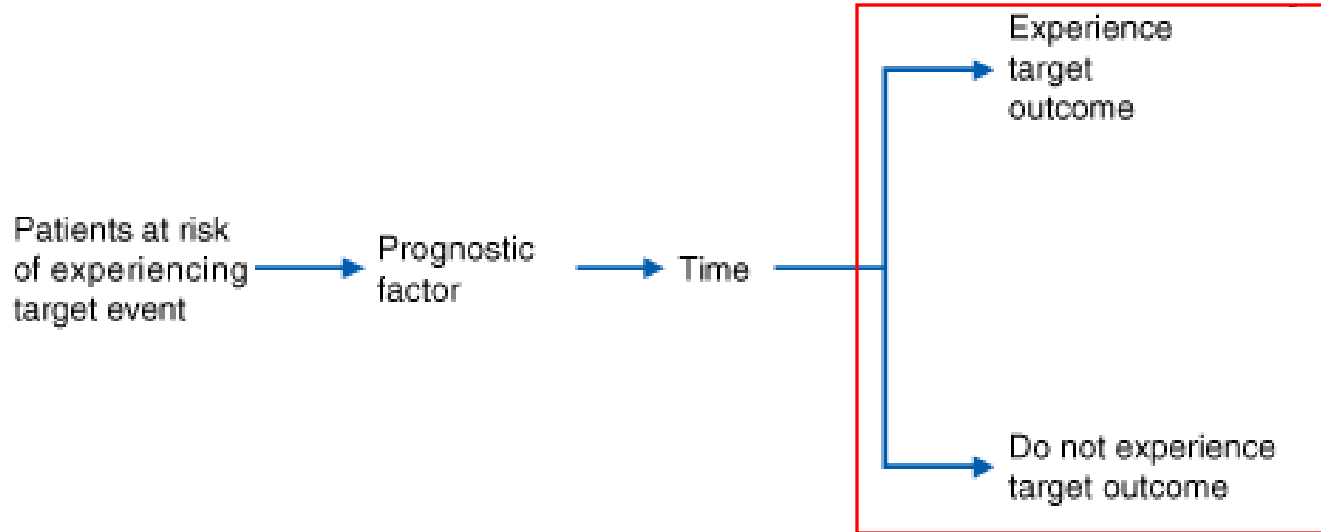
Diagnostic Test



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- Will the reproducibility of the test result and its interpretation be satisfactory in my setting?
 - Consider if test has intraobserver or interobserver variability?
- Are the results applicable to my patients?
- Will the results change my management?
 - What are the test and treatment thresholds for the health condition to be detected?
 - Are the test LRs high or low enough to shift posttest probability across a test or treatment threshold?
- Will the patients be better off as a result of the test?
 - Will patient care or outcomes differ for different test results?
- Will anticipated changes do more harm than good?

Prognosis Study



Source: Guyatt G, Rennie D, Meade MO, Cook DJ: *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice*, 2nd Edition: <http://www.jamaevidence.com>
Copyright © American Medical Association. All rights reserved.

- How large is the likelihood of the outcomes events in specified time period?
- How precise are the estimates of likelihood?
- Were the study patients and their management similar to those in my practice?
- Was the follow-up sufficiently long?
- Can I use the results in the management of patients in my practice?

Qualitative Study

- How evocative and thorough is the description? Does the data tell a story?
- Is the theory comprehensive, coherent, and appropriate?
- What are the concepts included in the theory?
- Does it make sense? Are concepts well developed?
- Does it fit with what we understand about the field? Are the results relevant?
- Do results in a report help each party (investigator, patient, reader) make sense of a social phenomena?
- Does this study teach me something that will help me understand my patients differently?
- Will it change my interactions in similar clinical situations in the future?

Surrogate Endpoints

- Is there a strong, independent, consistent relationship between the surrogate endpoint and the clinical endpoint?
- Strength of the study design: Relative Risk vs Odds Ratio, Cohort vs Case controlled study
- Is there evidence from other trials that improvement in the surrogate end point has consistently led to improvement in the target outcome?
- Is this population similar to the population I take care of?
- Are the likely treatment benefits worth the potential harm and costs? Overtreatment is not without risk

Screening

- What are the benefits?
 - Absolute risk reduction, Relative risk reduction, Number needed to treat
- What are the harms?
 - complications arising from investigation (screening test);
 - adverse effects of treatment;
 - adverse effects of labeling and early diagnosis;
 - anxiety generated by the investigations and treatment; and
 - costs and inconvenience incurred during investigations and treatment.
- How do benefits and harms compare in different people and with different screening strategies?
 - Risk of disease (Benefits change with risk of disease – eg age)
 - Screening interval – time sensitive
 - Test characteristic
- What is the effect of individuals' values and preferences?
 - Patient decision aids - Provide high quality balanced information about difficult decisions
 - Examples – Down Syndrome; Commercial drivers – strong economic incentives
- What is the effect of uncertainty associated with the evidence?