Peer review & Critical Appraisal

Dr. Kamran Yazdani, MD, MPH PhD

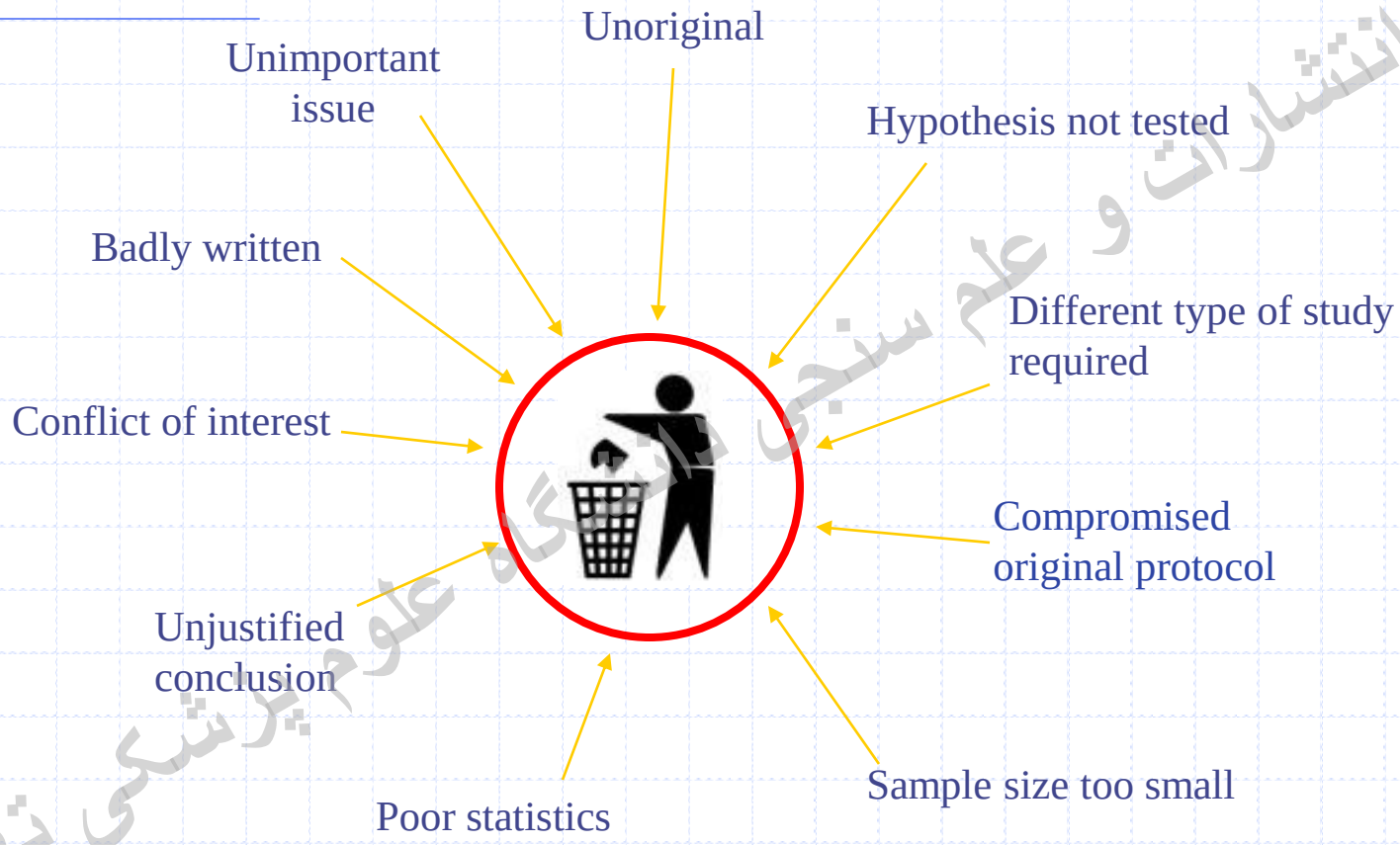
Vali Baigi, PhD Student

Special thanks to: Dr. Payam Kabiri, Dr. Leila Ghalichi, Dr. Gholamreza Khalili and Hamid Reza Tohidi

Office of Publications & Scientometrics

Tehran University of Medical Sciences

The science of 'trashing' a paper



Peer review

- Articles submitted to peer-reviewed journals (manuscripts) are reviewed by experts who advise the editor on whether they should be published and what changes are necessary.

Peer Review - Functions

- To Protect

  i) The author from publishing &

  ii) The subscriber from reading

Materials of insufficient quality

Editorial decision

An editorial committee may decide that a paper:

- Is acceptable for publication
- Is acceptable for publication following minor revisions
- Is acceptable for publication following major revision
- May be reconsidered for publication following major revisions
- May be considered for publication as a letter or a short report
- Is unacceptable for publication

Editorial decision

- Rejection rate: 15% (pay journals) to 60% (specialist journals) to more than 90% (NEJM, The Lancet)
- How long does it take? (Choice of journal)
  - BMJ: 70 days
  - JAMA: 117 days
  - Iranian journals?

Questions that journals ask

- Is the research question important?
- Is it interesting to our readers?
- Is it valid? A scientifically sound study.

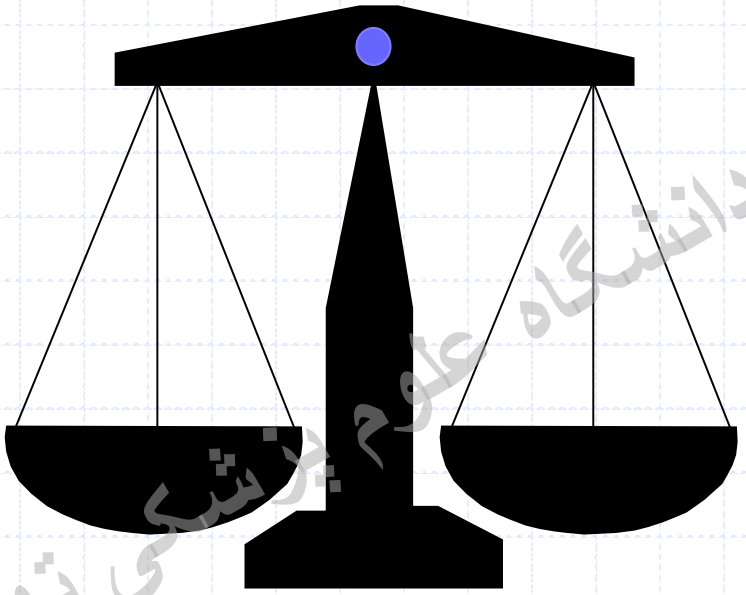
What editors and reviewers look for

- Short, clear, precise title
- Good abstract
- Good design and methods
- Appropriate statistics
- Simple tables and figures
- Comprehensive discussion
- Clear and fair conclusions
- Brevity, Balance, Logical organisation
- Follow instructions

Problems with peer review

- Slow
- Expensive
- A lottery
- Biased
- Easily abused
- Can't detect fraud

Critical appraisal



Critical appraisal is the use of explicit, transparent methods to assess the data in published research, applying the rules of evidence to factors such as internal validity, adherence to reporting standards, conclusions and generalizability.

applications of critical appraisal

- Decide how trustworthy a piece of research is ***(validity)***
- Determine what it is telling us ***(results)***
- Weigh up how useful the research will be ***(relevance)***
- a central part of the systematic review process

Critical Appraisal: Three preliminary questions

- **Why** was the study done and what hypothesis was being tested?
- **What** type of study was done?
- **Was** the study **design appropriate**?

Why was the study done?

i.e. what was the key research question/ what hypotheses were the author testing?

"null hypothesis"

To **investigate the association between residential radon and lung cancer** based on German living conditions, we conducted a case-control study in parts of western Germany.

Study designs:

Vali Baigi

Department of Epidemiology & Biostatistics
School of Public Health
Tehran University of Medical Sciences

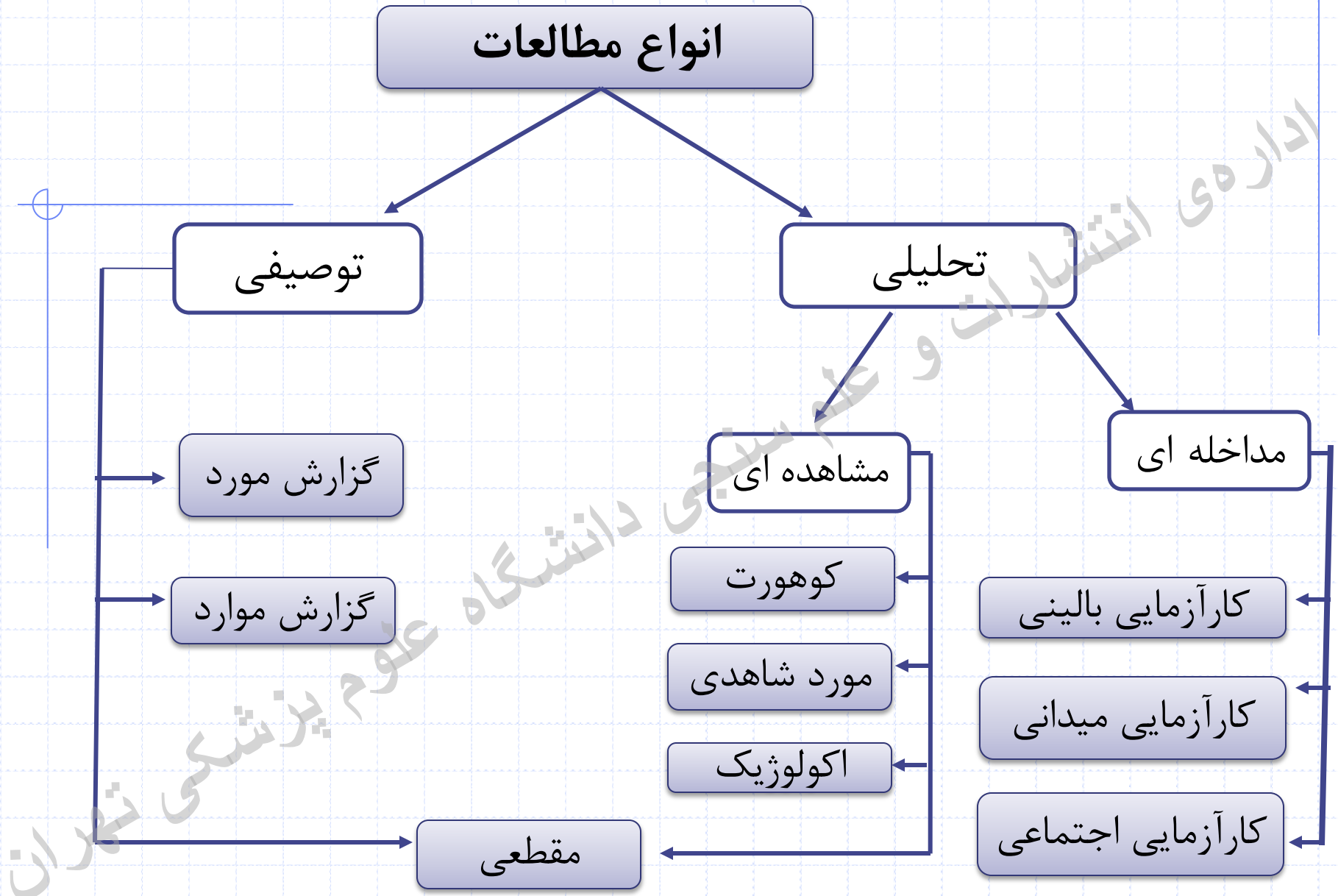
What type of study?

- **Qualitative**
- **Quantitative**

- **Primary** – these report research first hand.
- **Secondary** – summarise and draw conclusions from primary studies.

The Hierarchy of Evidence

1. Systematic reviews & meta-analyses
2. Randomised controlled trials
3. Cohort studies
4. Case-control studies
5. Cross sectional surveys
6. Case reports
7. Expert opinion
8. Anecdotal



Epidemiology

Definition?

تعریف اپیدمیولوژی:

John M Last:

"The study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems."

علم بررسی وفور، توزیع و عوامل تعیین کننده حالات مرتبط با سلامتی و حوادث در جوامع و بکار گیری این بررسی برای مبارزه با مشکلات بهداشتی.

اپیدمیولوژی

پاسخ به سوالات...

Who? When ? Where?

Why?

انواع مطالعات

- مطالعه‌های توصیفی
- مطالعه‌هایی هستند که پژوهشگر تنها وضعیت یک متغیر را بررسی کرده یا وضعیت چند متغیر را بدون در نظر گرفتن ارتباط آنها با یکدیگر بررسی می‌کند (شخص، زمان، مکان).
- تعیین بار بیماری، برآورد خدمات و نیروی انسانی
- ایجاد فرضیه (hypothesis generation)
- معمولا روی یک گروه ؟؟

انواع مطالعات

- **مطالعه‌های تحلیلی**
  - مطالعه‌هایی هستند که پژوهشگر به ارتباط بین دو یا چند متغیر پرداخته و هدف تعیین این ارتباط است.
  - روی حداقل دو گروه
  - آزمون فرض (hypothesis testing)

- مطالعات تحلیلی به دو سؤال زیر پاسخ می دهند:
- الف) آیا ارتباط بین علت و معلول وجود دارد؟
- ب) اگر ارتباطی وجود دارد آیا این ارتباط علیتی است؟

مهم ترین انواع مطالعات توصیفی

- گزارش مورد (بیمار) Case Report
- مجموعه موارد Case Series
- مقطعی Cross-sectional

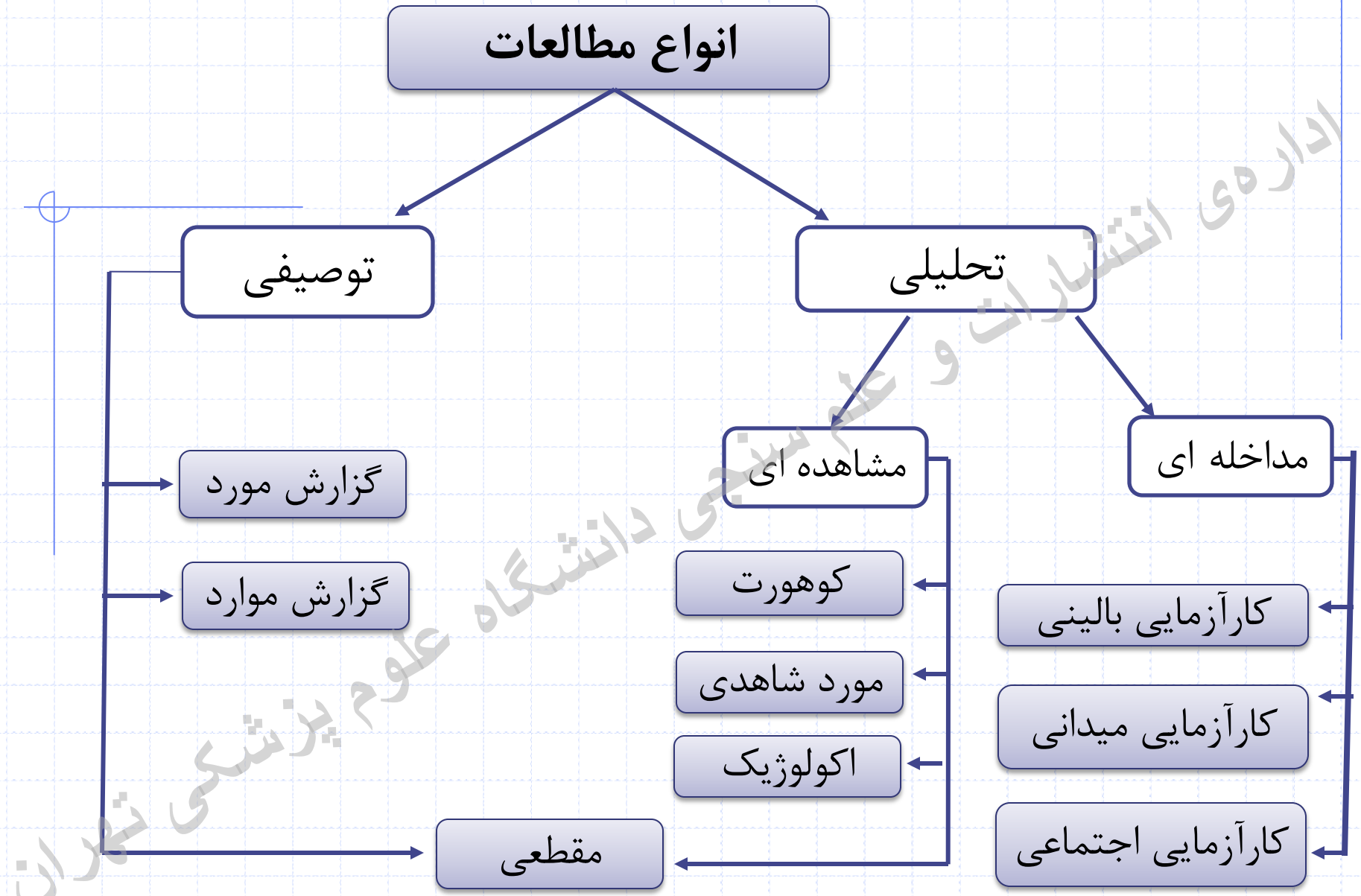
تقسیم‌بندی مطالعه‌های تحلیلی

- **مطالعه‌های مشاهده‌ای**
  - مطالعه‌هایی هستند که در آن پژوهش‌گر هیچ نقشی در وجود و مقدار متغیرهای مستقل و مخدوش‌کننده در بین واحدهای پژوهش ندارد.

- **مطالعه‌های مداخله‌ای**
  - مطالعه‌های هستند که پژوهش‌گر حداقل یک متغیر مستقل (مواجهه) را خود، تعیین می‌کند.

انواع مطالعات تحلیلی

- **مطالعات تحلیلی مشاهده ای**
  - **مقطعی** Cross-sectional
  - **اکولوژیک** Ecological
  - **مطالعه کوهورت** Cohort studies
  - **مطالعه مورد شاهدی** Case-control
- **مطالعات تحلیلی مداخله ای**
  - **مطالعات کارآزمایی بالینی** Clinical Trial
  - **مطالعات تجربی** Experimental studies



انتخاب جمعیت مرجع و جمعیت مورد مطالعه

- جمعیت مرجع یا جامعه هدف (Reference or Target population)
  - جامعه ای که انتظار می رود مداخله مورد نظر در مطالعه تجربی برای آنها منافعی در پی داشته باشد و نتایج مطالعه به آنها تعمیم داده می شود.
    - مثال: بیماران مبتلا به پرفشاری خون خفیف
- جمعیت مورد مطالعه یا مداخله (Study or Experimental population)

تعیین معیارهای انتخاب یا واجد شرایط بودن افراد (eligibility criteria)

- معیارهای ورود (inclusion criteria)
- معیارهای خروج (exclusion criteria)

انتخاب افراد مورد مطالعه (ادامه)

- **معیارهای ورود (inclusion criteria):**
  - مثال: بیماران مراجعه کننده به درمانگاه های داخلی بیمارستان که:

۱- بیمار مبتلا به پرفشاری خون خفیف (براساس تعریف WHO) باشد.

۲- سن فرد ۲۵ تا ۴۹ سال باشد.

۳- بیمار ساکن تهران باشد.

- **معیارهای خروج (exclusion criteria):**
  - مثال: بیماران فوق در صورت داشتن هر یک از خصوصیات زیر از مطالعه خارج می شوند:

۱- بیمار مبتلا به پرفشاری خون ثانویه باشد.

۲- بیمار سیگاری یا دیابتی یا چاق (BMI≥30) باشد.

۳- بیمار دچار بیماری ایسکمیک قلب، نارسایی کلیه یا هرنوع بیماری باشد که در اثر عدم کنترل پرفشاری خون تشدید شود.

مطالعات مقطعی

Cross Sectional Studies

مطالعات مقطعی (Cross-sectional studies)

- این مطالعات معمولا بر روی نمونه‌ای تصادفی از گروه تعریف شده‌ی اصلی انجام می‌گیرد.
- بررسی افراد یک جمعیت (هر فرد فقط یک بار بررسی می شود)
- متداولتر از سایر مطالعات است
- مطالعه مقطعی که به عنوان مطالعه «شیوع» نیز خوانده می شود، معمولاً شامل یک نمونه گیری تصادفی از جمعیت هدف است.
- در مرحله بعد فراوانی بیماری و وضعیت مواجهه های فعلی یا قبلی افراد و سایر متغیرهای مورد علاقه محقق در این نمونه بررسی می شود.

اهمیت نمونه گیری در مطالعات مقطعی

Cross Sectional Studies (contd)

- نمونه گیری بدون اطلاع از وضعیت بیماری یا مواجهه انجام می گیرد.
- نمونه گیری در یک نقطه زمانی خاص انجام می گیرد.
- این مطالعات توانایی تعیین شیوع یک پدیده را در جامعه دارند.

مزایای مطالعات مقطعی

- **به عنوان یک مطالعه توصیفی طرح مناسب برای تولید فرضیه**

  (معمولاً اولین قدم در بررسی یک موضوع خاص هستند.)

- **مناسب برای برآورد پارامترهای مورد نظر در جامعه (کمی و کیفی)**

سیاست گذاری و برنامه ریزی

- **توانایی به کاربردن برای چندین مواجهه مختلف یا چندین پیامد خاص**

محدودیتهای مطالعات مقطعی

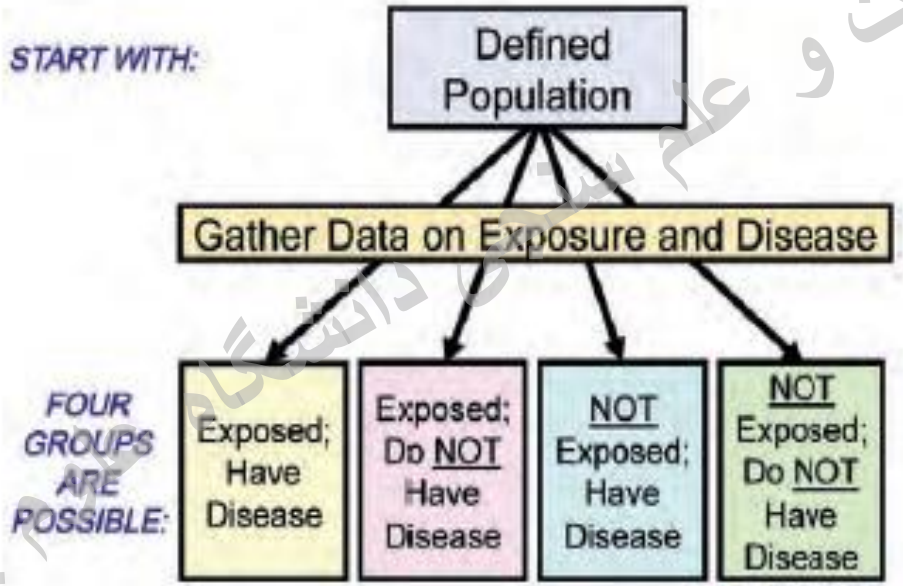
- عدم توانایی اندازه گیری بروز
- پایین بودن **level of evidence**

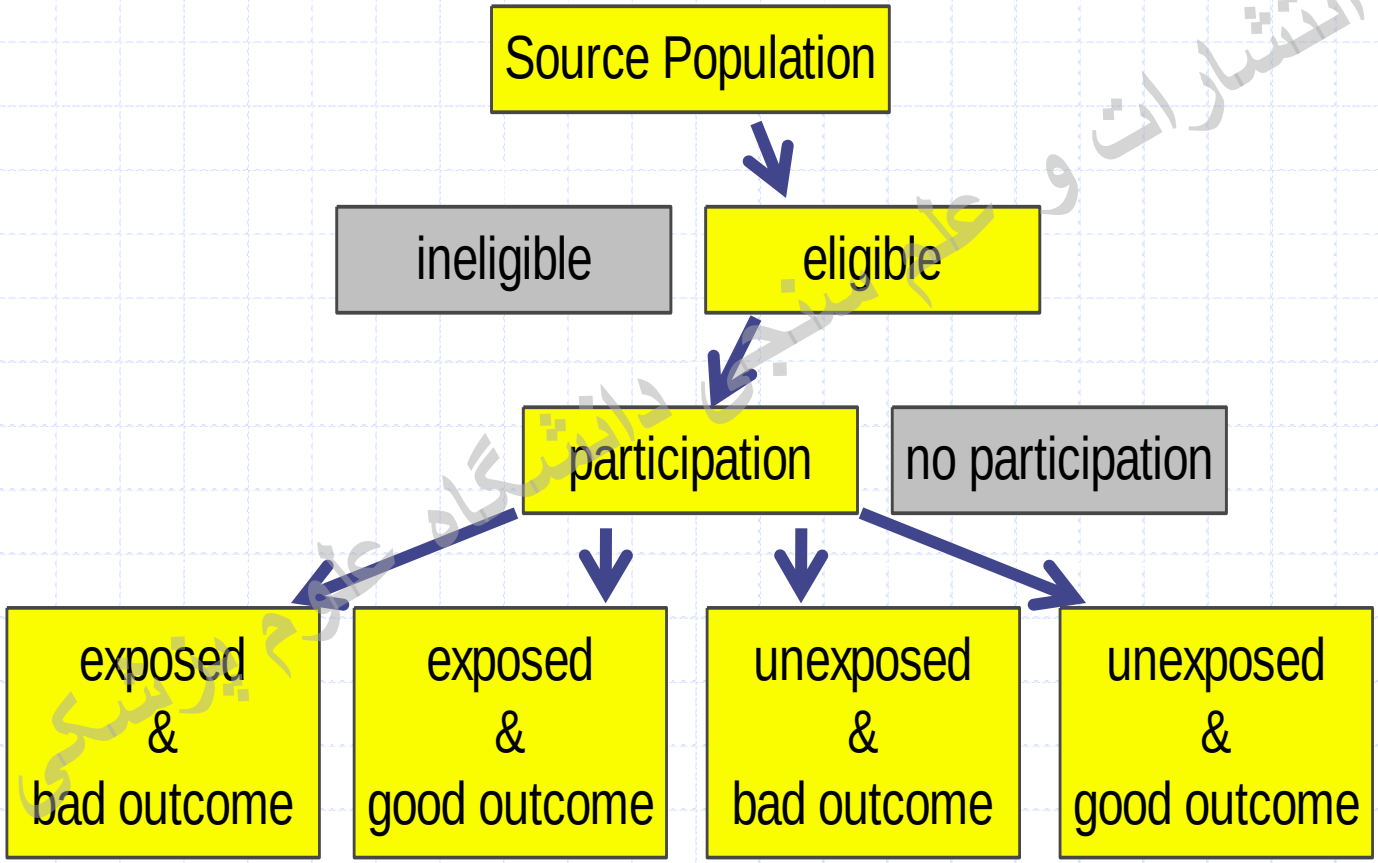
-خطاها و مخدوش کننده ها

-مشکل تفسیر رابطه زمانی

مشکل تفسیر رابطه زمانی در مطالعات مقطعی

- ✓ رابطه بین وضعیت اقتصادی ـاجتماعی و افسردگی
- ✓ رابطه بین فشار خون بالا و جنسیت





Cross Sectional Studies

هدف: بررسی رابطه سطح اقتصادی- اجتماعی (SES) با افسردگی (Depression) در افراد ۱۸ تا ۲۲ ساله شهر تهران

Cross Sectional Studies

	Disease +	Disease -	Total
exposure	a	b	a + b
Non-exposure	c	d	c + d
Total	a + c	b + d	N

Cross Sectional Studies

وضعیت اقتصادی اجتماعی	افسردگی: بیمار	افسردگی: سالم	جمع کل
خوب	۳	۸۷	۹۰
بد	۱۴	۷۵	۸۹
جمع کل	۱۷	۱۶۲	۱۷۹

Special considerations in this study:

- **Choosing a representative sample** (Sampling strategy)
- **Sample size (precision)**
- **Data collection**
- **Potential bias in cross-sectional studies**

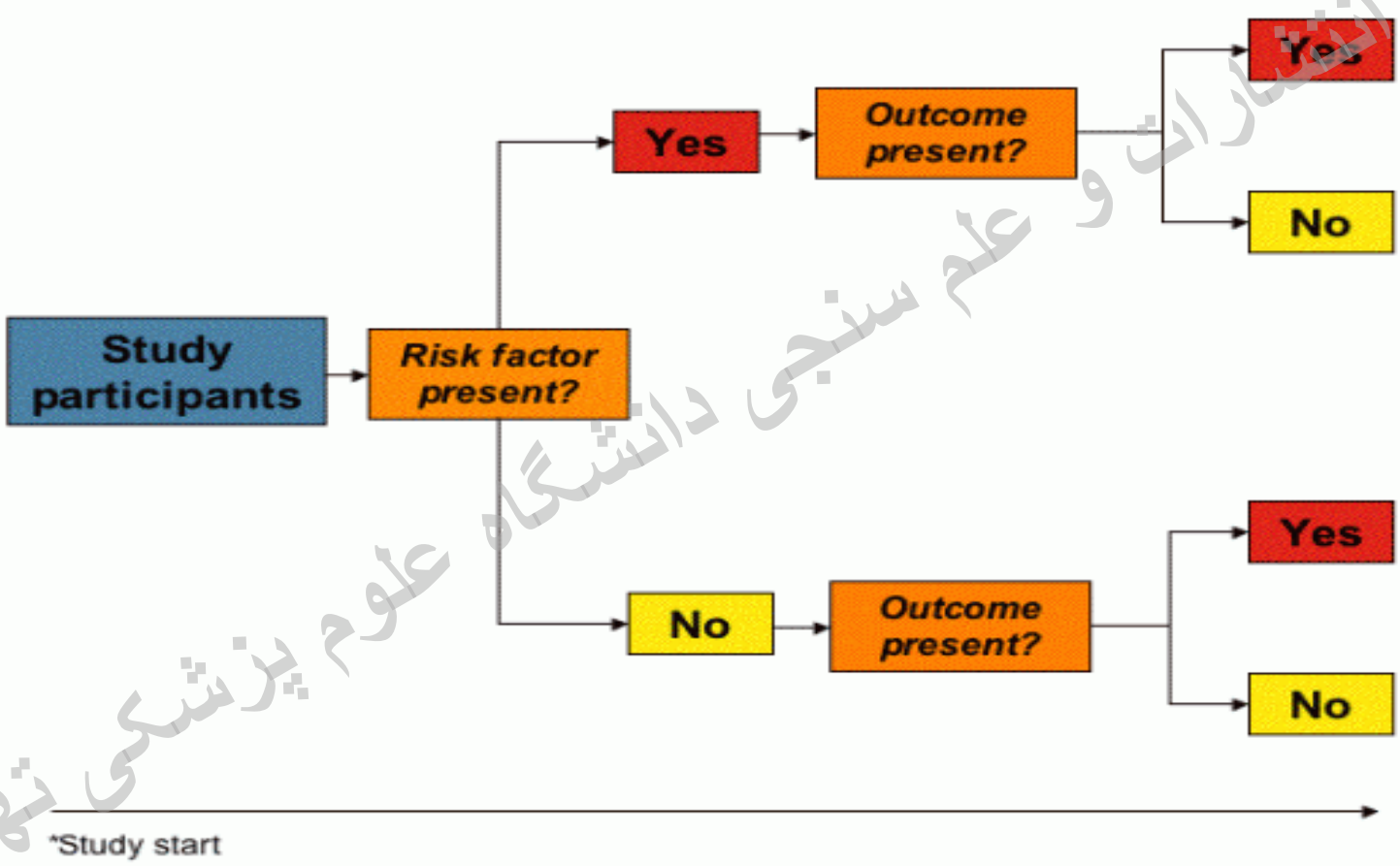
Non-response is a particular problem affecting cross-sectional studies and can result in bias of the measures of outcome. This is a particular problem when the characteristics of non-responders differ from responders.

مشخصات مطالعات همگروهی

- هم گروهها (مواجهه‌دار و غیر مواجهه‌دار) قبل از ظهور بیماری مشخص می‌شوند.
- سمت و جهت مطالعه از مواجهه به سوی بیماری می‌باشد.
- گروههای مورد مطالعه را در طی دوره زمانی مشخص از نظر ظهور بیماری پیگیری (Follow)میکنیم.
- اندازه گیری بروز در گروه مواجهه دار و عدم مواجهه

- مناسب تر برای مواجهه های نادر
- هزینه بالا و زمان نسبتا طولانی به دلیل نیاز به پیگیری افراد مطالعه

مطالعات همگروهی



COHORT STUDY

	Disease +	Disease -	Total
exposure	a	b	a + b
Non-exposure	c	d	c + d
Total	a + c	b + d	N

special considerations in cohort study:

- Selection of study groups

They are identical except for their exposure status. They should be at risk.

- Measuring exposure
- Measuring outcome
- Methods of follow up
- Major source of potential bias: losses to follow up

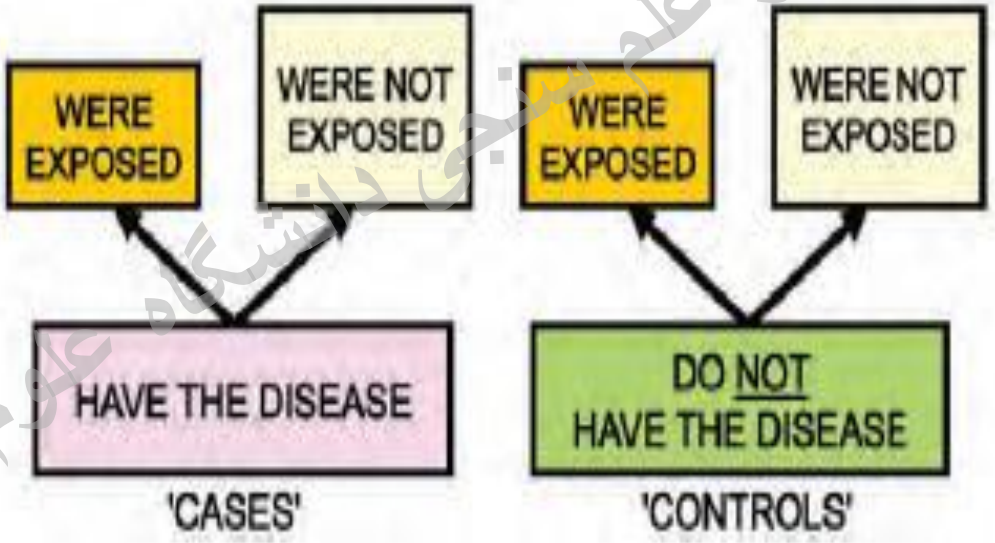
مطالعات مورد شاهدی

- سمت و جهت مطالعه از بیماری به سوی مواجهه می‌باشد.
- از گروه شاهد برای رد یا قبول فرضیه استفاده می‌شود.
- مناسب تر برای بیماری های نادر
- نسبتا ارزان قیمت

مهمترین نقطه ضعف (پاشنه آشیل) مطالعات مورد شاهدی

انتخاب گروه شاهد

مطالعات مورد شاهدی



Case-Control STUDY

	Disease +	Disease -	Total
exposure	a	b	a + b
Non-exposure	c	d	c + d
Total	a + c	b + d	N

Special considerations in this study:

- Case definition
- Source of cases
- Selection of cases (incident or prevalent cases).
- Selection of controls
- Measuring exposure status (it should be the same in both groups)
- prone to recall and observer bias

Epidemiologic Study Designs

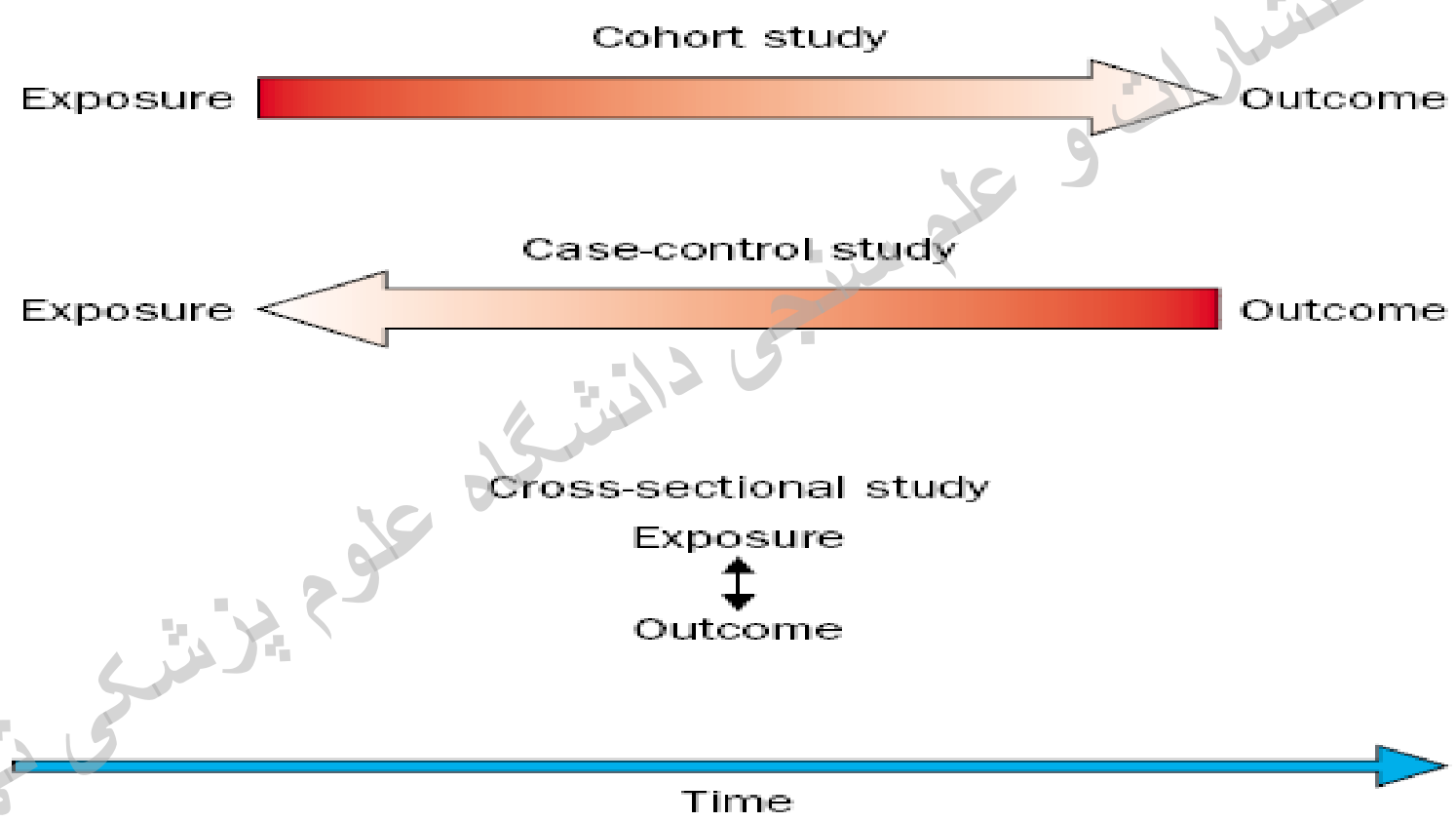


Figure 2: **Schematic diagram showing temporal direction of three study designs**

Grimes & Schulz, 2002 (www)

تفاوت های اصلی بین مطالعات مورد شاهدی و هم گروهی (۱)

مطالعات مورد شاهدی	مطالعات هم گروهی
۱- از معلول به علت می رسد	از علت به معلول می رسد
۲- مناسب تر برای بیماری های نادر	مناسب تر برای مواجهه های نادر
۳-محاسبه شیوع مواجهه در افراد بیمار و افراد بدون آن بیماری	محاسبه بروز بیماری در افراد مواجهه یافته و افراد بدون مواجهه
۵- نیاز به حجم نمونه کم	حجم نمونه بالا
۶- نسبتا ارزان و سریع	نسبتا پرهزینه و وقت گیر
۷- میتوان رابطه چندین مواجهه با یک بیماری (پیامد) را بررسی کرد.	میتوان رابطه چندین پیامد با یک مواجهه را بررسی کرد
۸- معمولاً فقط برآوردی از خطر نسبی را به صورت نسبت شانس به دست می دهد.	خطر نسبی و همچنین خطر قابل انتساب (خطر منتسب) را بطور مستقیم محاسبه می کند.

تعریف کارآزمایی بالینی

Randomized Clinical Trial (RCT)

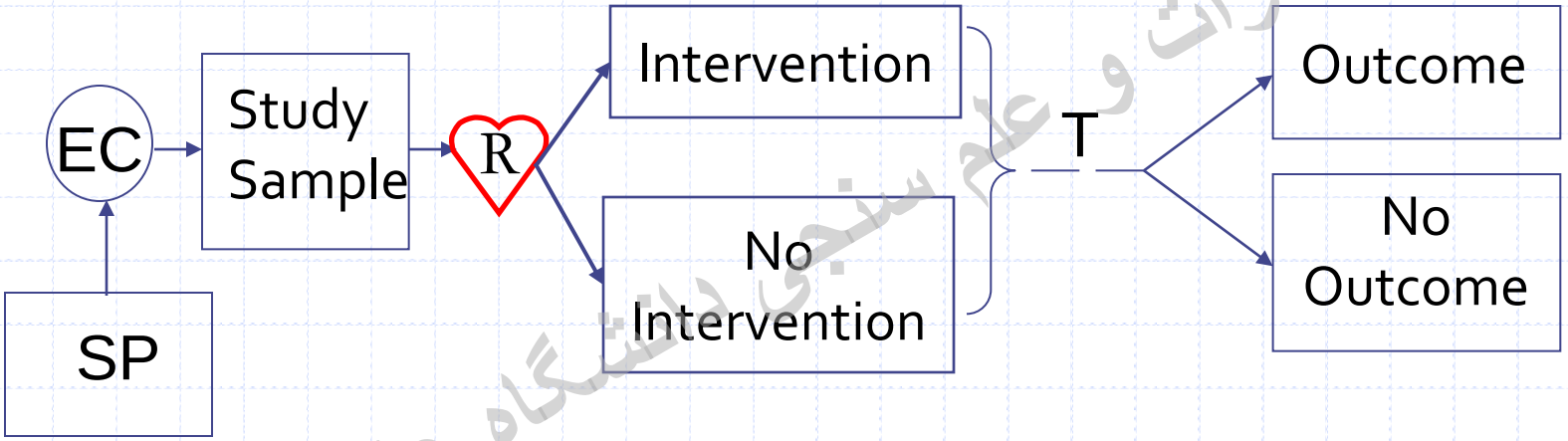
- کارآزمایی بالینی مطالعه ای است آینده نگر که برای مقایسه اثرات و ارزش یک مداخله (یا مداخله ها) در برابر شاهد در نمونه های انسانی انجام می شود.

هدف کارآزمایی بالینی

- ارزیابی کارایی (efficacy) و اثربخشی (effectiveness) یک مداخله یا داروی جدید

- مداخله: روش درمانی، روش تشخیصی (غربالگری)، روش آموزشی، روش پیشگیری....

طرح کلی کارآزمایی بالینی



SP = Study Population
EC = Eligibility Criteria
R = Randomize intervention
T = Elapsed time

Who gets which treatment?

- To conduct a good experiment, "**treatment assignments**" must be "**random**."

- "**Random**" means everybody has an equal chance of getting a treatment.

Random Selection vs Random Assignment

- Random selection is how you draw the sample of people for your study from a population

- Random assignment is how you assign the sample that you draw to different groups or treatments in your study

Advantages of experiments

- Randomization **should** make the two populations similar, on average, with respect to everything except the treatment.

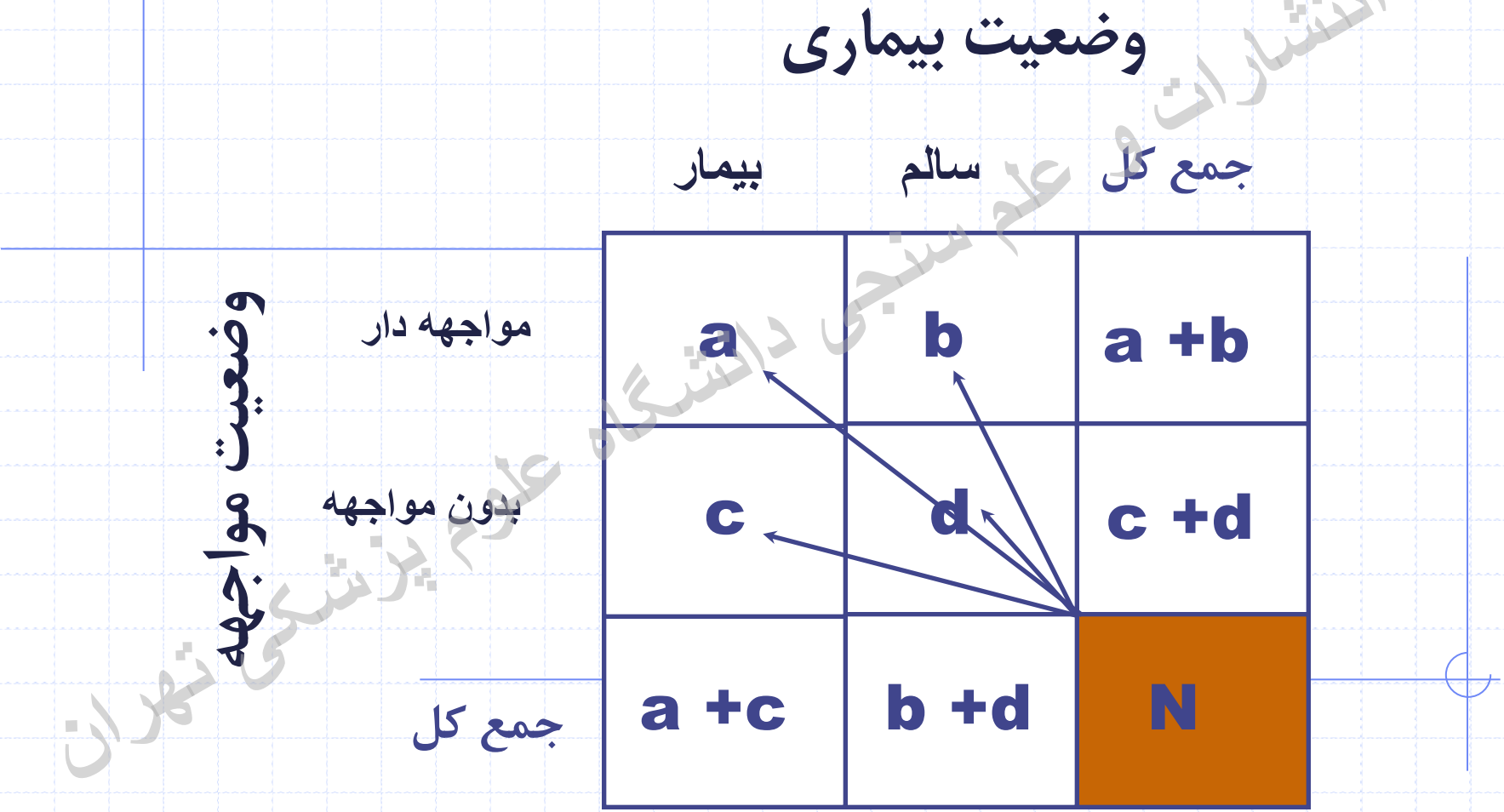
- So if outcomes are different for the two populations, can conclude that it is the treatment that **caused** it.

special considerations in RCTs:

- Method of Randomization
- Allocation concealment
- Blinding (Masking)
- Ethical issues
- RCT registration
- Analysis method (ITT, per Protocol or as treated)

Review of what we learnt

Cross Sectional Studies



COHORT STUDY

	Disease +	Disease -	Total
exposure	a	b	a + b
Non-exposure	c	d	c + d
Total	a + c	b + d	N

Case-Control Study

	Disease +	Disease -	Total
exposure	a	b	a + b
Non-exposure	c	d	c + d
Total	a + c	b + d	N

Any question?

practice 1

Measures of Occurrence & Association

Dr. Kamran Yazdani, MD, MPH PhD

Different Types of Fractions

- Ratio
- Proportion
- Rate

- **Ratio**: A fraction in which the numerator is not part of the denominator.

e.g. Fetal Death Ratio:

fetal deaths/live births

Fetal deaths are not included among live births, by definition.

- **Proportion**: A fraction in which the numerator is part of the denominator.

e.g. Fetal Death Rate:

fetal deaths/all births

All births includes both live births and fetal deaths.

- **Rate**: A proportion in which change over time is considered

Measures of occurrence:

- Prevalence (existing cases)

$$\text{Point Prevalence} = \frac{\text{No. of cases in a defined population at one point in time}}{\text{No. of persons in a defined population at the same point in time}}$$

Measures of occurrence:

- **Incidence (new cases)**
- Cumulative Incidence (Risk)

$$\text{Incidence Risk} = \frac{\text{Number of new cases of disease in a specified period of time}}{\text{Number of disease-free persons at the beginning of that time period}}$$

- Incidence density (Rate)

$$\text{Incidence rate} = \frac{\text{Number of new cases of disease in a given time period}}{\text{Total person-time at risk during the follow-up period}}$$

Measures of Association

- **Ratios:**
  - ✓ Risk Ratio (**Relative Risk**)
  - ✓ Rate Ratio (Relative Rate)
  - ✓ **Odds Ratio** (Relative Odds)
- **Differences:**
  - ✓ Risk difference (Attributable Risk)

وضعیت بیماری

وضعیت مواجهه	بیمار	سالم	جمع کل
مواجهه دار	a	b	a +b
بدون مواجهه	c	d	c +d
جمع کل	a +c	b +d	N

Relative Risk (RR)

		Disease +	Disease −	Total
Exposed	+	a	b	a+b
	−	c	d	c+d

$$\text{Relative Risk} = \frac{\text{Incidence in Exposed}}{\text{Incidence in Nonexposed}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Positive Association (Risk Factor)	RR > ۱ : Risk in Exposed > Risk in Nonexposed
Negative Association (Protective Effect)	RR < ۱ : Risk in Exposed < Risk in Nonexposed
No Association (Null Effect)	RR = ۱ : Risk in Exposed = Risk in Nonexposed

Odds

- Example: horse race

$$\text{Probability of winning } (P) = 60\%$$

$$\text{Probability of losing } (1 - P) = 40\%$$

$$\text{Odds of winning} = \frac{\text{Probability of winning}}{\text{Probability of losing}} = \frac{P}{1 - P} = \frac{60\%}{40\%} = 1.5$$

Odds Ratio (OR)

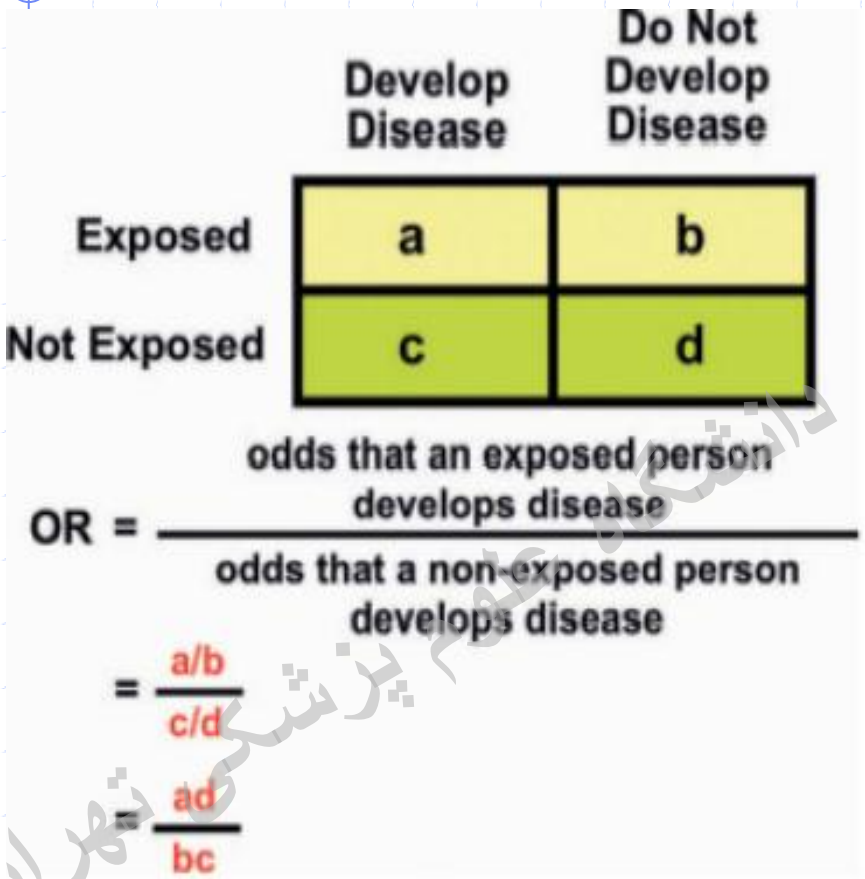
		Diseased	Nondiseased	Total
Exposed	+	a	b	a+b
	−	c	d	c+d

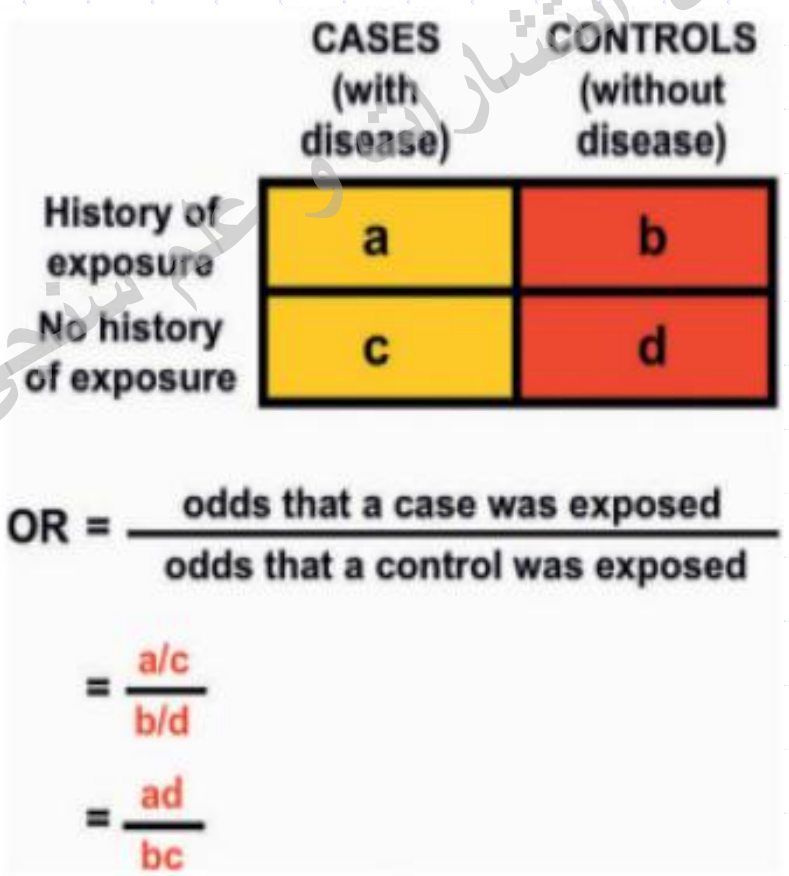
$$\text{Odds of Disease in Exposed} = \frac{P}{1-P} = \frac{\frac{a}{a+b}}{\frac{b}{a+b}} = \frac{a}{b}$$

$$\text{Odds of Disease in Nonexposed} = \frac{P}{1-P} = \frac{\frac{c}{c+d}}{\frac{d}{c+d}} = \frac{c}{d}$$

$$\text{Odds Ratio (OR)} = \frac{\text{Odds of Disease in Exposed}}{\text{Odds of Disease in Nonexposed}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

A:Odds ratio (OR) in a cohort study
B: Odds ratio (OR) in a case-control study





Relative Risk versus Attributable Risk

Relative Risk

- is a measure of the strength of the association.
- is important to establish etiologic relationship.

Attributable Risk

- is the potential for prevention if the exposure could be eliminated.
- is important in clinical practice and public health.

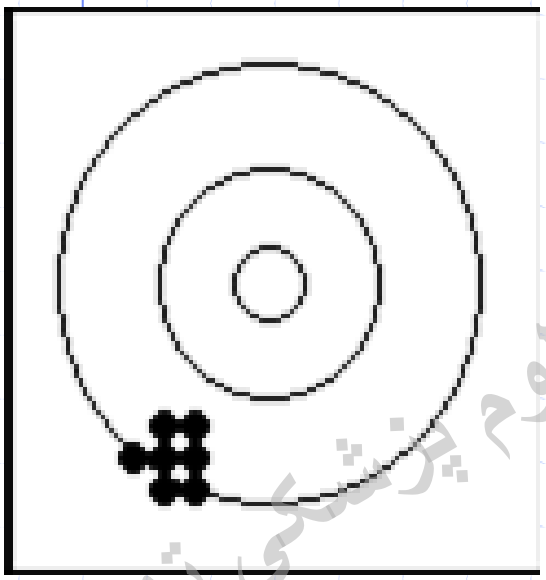
Practice 2

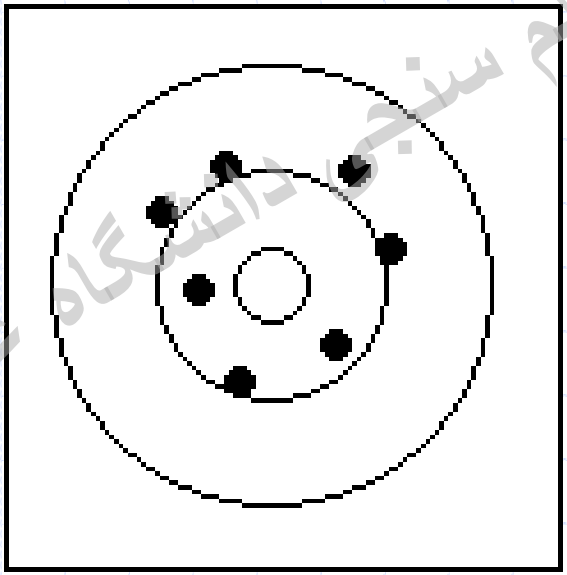
Errors in Research

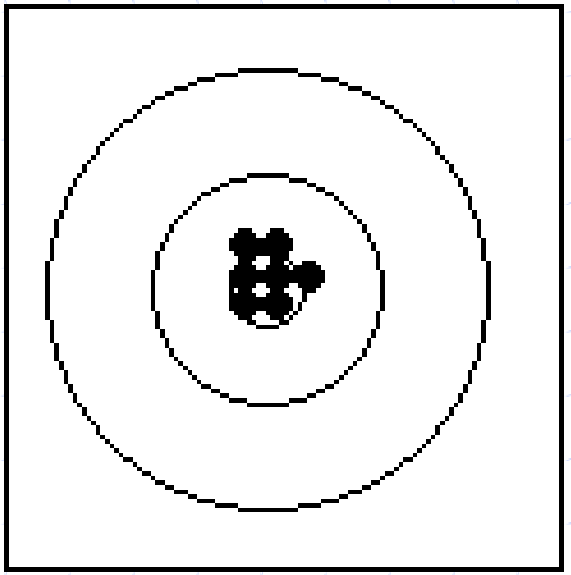
Vali Baigi

Associations may be due to

- **Chance (random error)**
  - statistics are used to reduce it by appropriate design of the study
  - statistics are used to estimate the probability that the observed results are due to chance
- **Bias (Systematic error)**
  - must be considered in the design of the study
- **Confounding**
  - can be dealt with during both the design and the analysis of the study
- **True association**







مثال: فردی که می‌دانیم ۷۰ کیلوگرم وزن دارد را ۵ بار با چهار ترازوی مختلف وزن می‌کنیم . . . نظر شما در هر مورد چیست؟

	ترازوی چهارم	ترازوی سوم	ترازوی دوم	ترازوی اول
وزن در سنجش اول	۶۹	۷۱	۶۹	۷۰
وزن در سنجش دوم	۷۲	۷۱	۶۸	۷۰
وزن در سنجش سوم	۷۱	۷۱	۷۱	۷۰
وزن در سنجش چهارم	۷۲	۷۱	۷۲	۷۰
وزن در سنجش پنجم	۷۱	۷۱	۷۲	۷۰
میانگین	۷۱	۷۱	۷۰	۷۰

خطای تصادفی

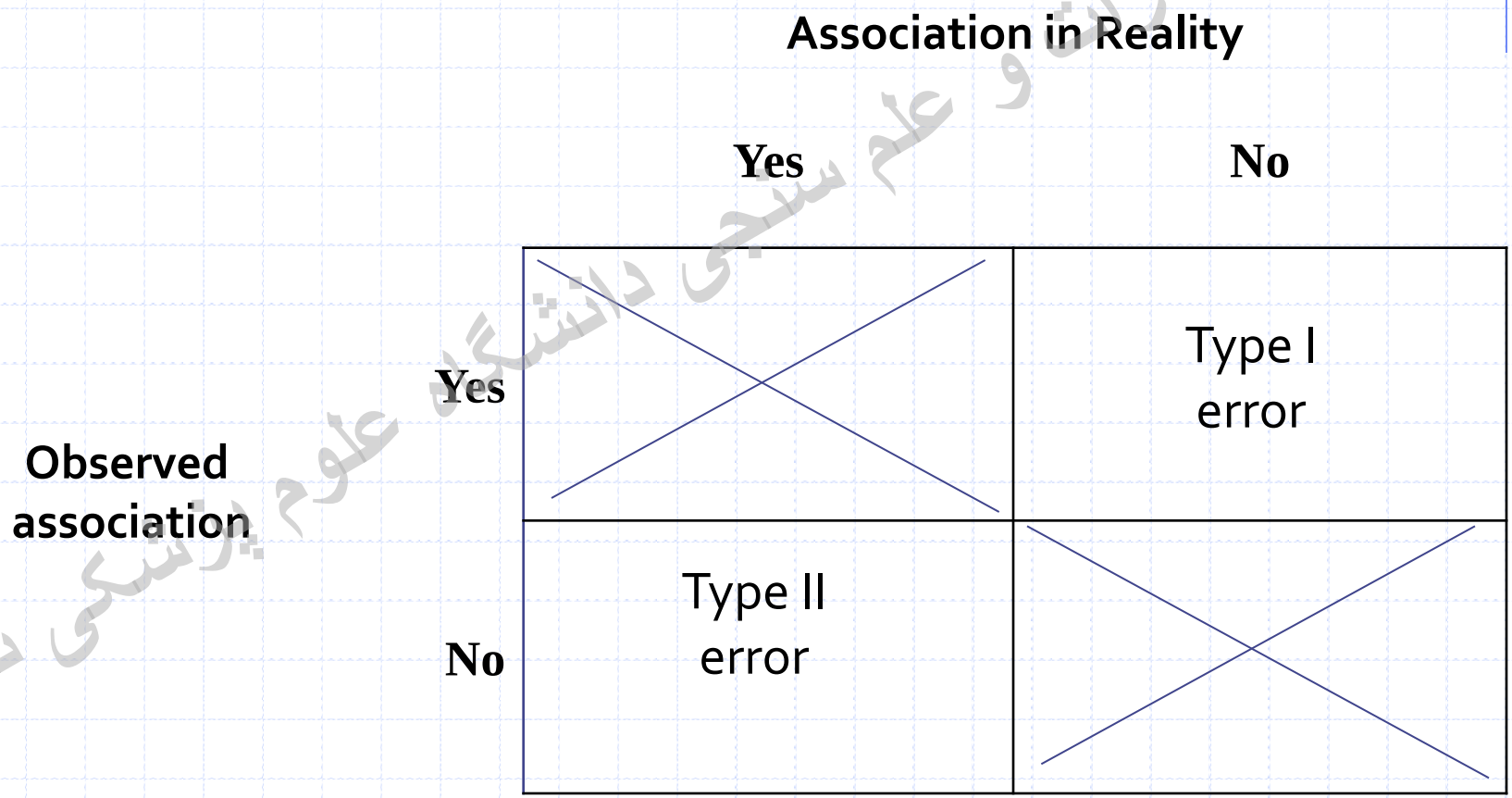
- ناشی از تغییرات معمول در مقدار اندازه‌گیری شده.
- ناشی از تغییرات حاصل از عدم دقت وسیله‌ی اندازه‌گیری.

Dealing with chance error

- **During design of study**
  - Sample size
  - Power
- **During analysis (Statistical measures of chance)**
  - Test of statistical significance (P value)
  - Confidence intervals

Statistical measures of chance I

(Test of statistical significance)

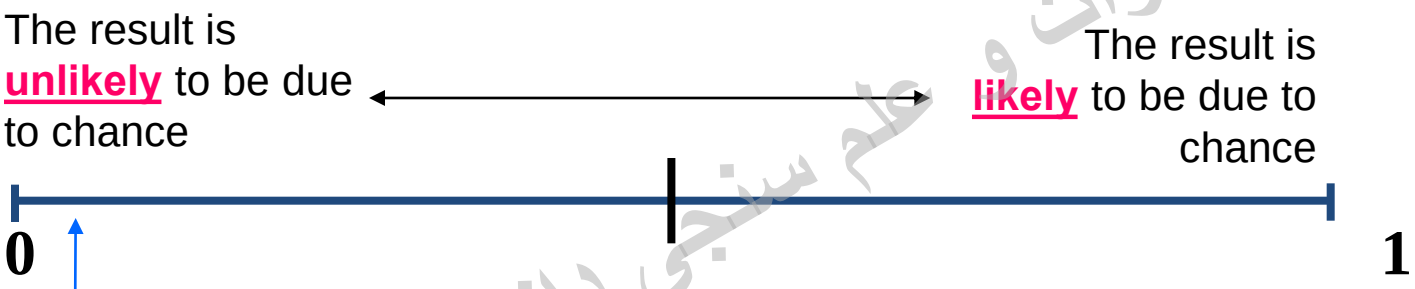


P-value

- the probability the observed results occurred by chance
- the probability that an effect at least as extreme as that observed could have occurred by chance alone, given there is truly no relationship between exposure and disease (Ho)
- statistically non-significant results are not necessarily attributable to chance due to small sample size

The p-value in a nutshell

Could the result have occurred by chance?



p < 0.05
a statistically significant result

p > 0.05
not a statistically significant result

p = 0.05

$\frac{1}{20}$ or 1 in 20 result fairly unlikely to be due to chance

P-value

0.00001

Clinical Importance
VS
Statistical Significance

Statistical Power

- Power = 1 – type II error
- Power = 1 - ß

Statistical measures of chance II (Confidence intervals)

Answer: Confidence Interval

- **Definition:** A range of values for a variable of interest constructed so that this range has a specified probability of including the true value of the variable for the population
- **Characteristics:**
  - a measure of the precision (stability) of an observed effect
  - the range within which the true magnitude of effect lies with a particular degree of certainty
  - 95% C.I. means that true estimate of effect (mean, risk, rate) lies within 2 standard errors of the population mean 95 times out of 100
  - Confidence intervals get smaller (i.e. more precise or more certain) if the underlying data have less variation/scatter
  - Confidence intervals get smaller if there are more people in your sample

How to Estimate CI?

- Standard Error (SE)
- 95% CI = statistic ± 1.96 SE
- Example:
  - 95% CI of mean = sample mean ± 1.96 SE
  - $SE = \frac{SD}{\sqrt{n}}$

Question?

- 20 out of 100 participants: 20%
- 200 out of 1000 participants: 20%
- 2000 out of 10000 participants: 20%
- What is the difference?

95% Confidence Interval (95% CI)

- 20 out of 100 participants: 20%

  95% CI: 12 to 28

- 80 out of 400 participants: 20%

  95% CI: 16 to 24

- 2000 out of 10000 participants: 20%

  95% CI: 19.2 to 20.8

How to Estimate CI? (example)

- A sample of 100 participants
- Mean of their age 25 years
- SD of age: 10
- CIM?

- CIM = $25 \pm 1.96 * 10 / \sqrt{100}$
- CIM ~ from 23 to 27

How to Estimate CI for proportion?

- Standard Error (SE)
- 95% CI = statistic ± 1.96 SE

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

Associations may be due to

- **Chance (random error)**
  - statistics are used to reduce it by appropriate design of the study
  - statistics are used to estimate the probability that the observed results are due to chance
- **Bias (Systematic error)**
  - must be considered in the design of the study
- **Confounding**
  - can be dealt with during both the design and the analysis of the study
- **True association**

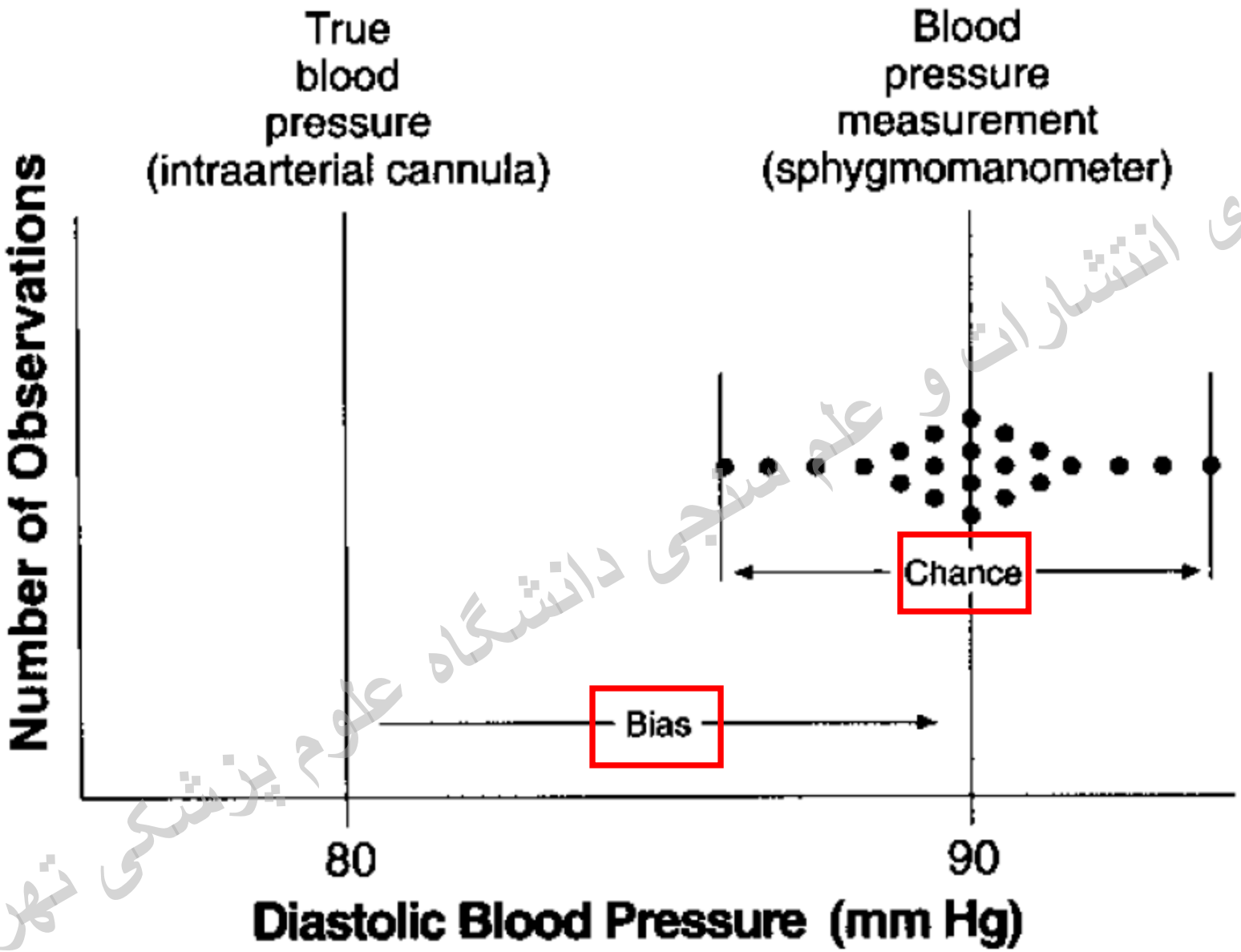


Figure 1.2. Relationship between bias and chance: Blood pressure measurements by intraarterial cannula and sphygmomanometer.

Bias:

-**تعریف:** هر خطای سیستماتیکی که موجب تخمین نادرستی از رابطه بین عامل خطر و پیامد می شود.

- ویژگیها:

- این نوع خطا دارای جهت یا الگوی خاصی است.

- به حجم نمونه بستگی ندارد.

- میانگین در این نوع خطا با میانگین واقعی برابر نیست.

Types of Bias

- Selection bias – identification of individual subjects for inclusion in study on the basis of either exposure or disease status depends in some way on the other axis of interest

- Observation (information) bias – results from systematic differences in the way data on exposure or outcome are obtained from the various study groups

Control of Bias

- Can only be prevented and controlled during the design and conduct of a study:
  - Careful planning of measurements
  - Formal assessments of validity
  - Regular calibration of instruments
  - Training of data collection personnel
  - Blinding

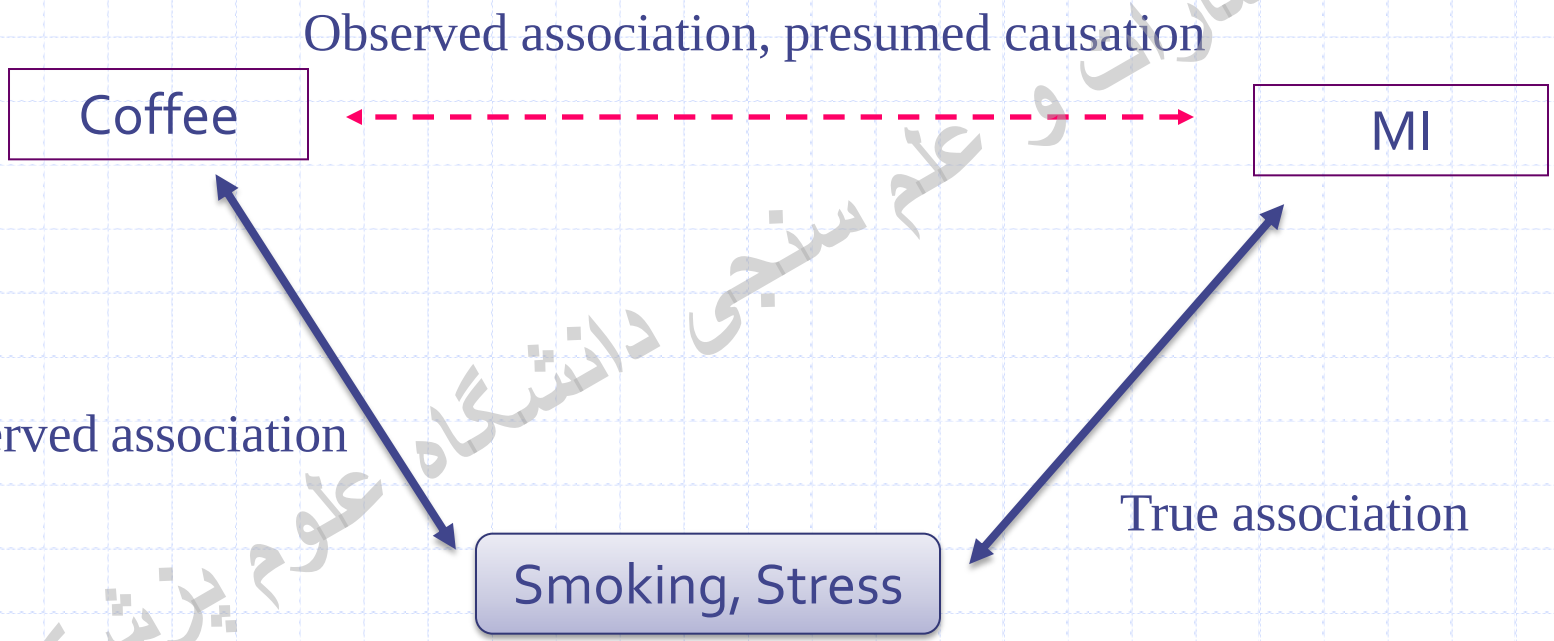
Associations may be due to

- **Chance (random error)**
  - statistics are used to reduce it by appropriate design of the study
  - statistics are used to estimate the probability that the observed results are due to chance
- **Bias (Systematic error)**
  - must be considered in the design of the study
- **Confounding**
  - can be dealt with during both the design and the analysis of the study
- **True association**

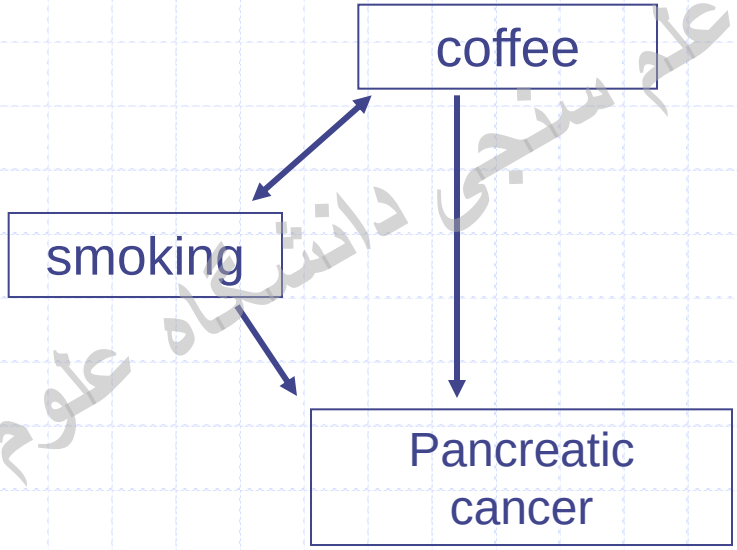
Confounding

- Confounding results when the effect of an exposure on the disease (or outcome) is distorted because of the association of exposure with other factor(s) that influence the outcome under study.

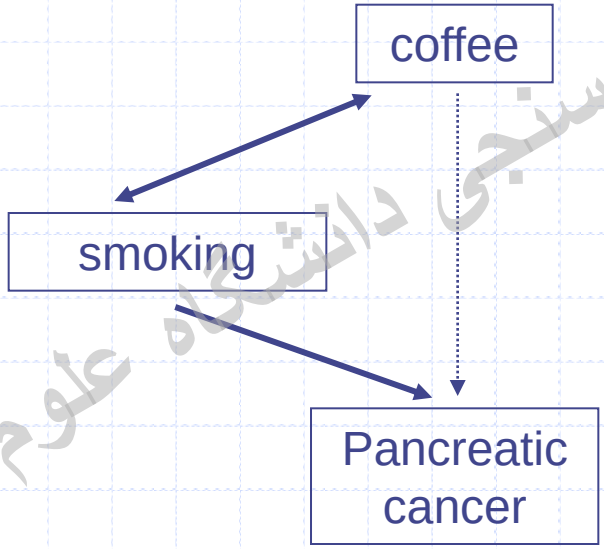
Confounding



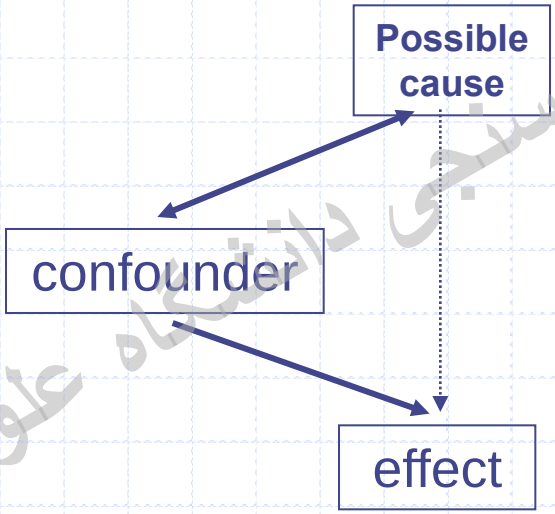
Confounding



Confounding



Confounding



Control of Confounding

- **During design of study**
  - Restriction
  - Matching
  - Randomization
- **During analysis**
  - Stratified analysis
  - Multivariate analysis (regression)

Associations may be due to

- **Chance (random error)**
  - statistics are used to reduce it by appropriate design of the study
  - statistics are used to estimate the probability that the observed results are due to chance
- **Bias (Systematic error)**
  - must be considered in the design of the study
- **Confounding**
  - can be dealt with during both the design and the analysis of the study
- **True association**

DETERMINATION OF CAUSATION

- The general QUESTION: Is there a cause and effect relationship between the presence of factor X and the development of disease Y?

Nature of Evidence:

1. Replication of Findings
   - consistent in populations

2. Strength of Association
   - significant high risk

3. Temporal Sequence
   - exposure precede disease

Nature of Evidence:

4. Dose-Response
   - higher dose exposure, higher risk

5. Biologic Credibility
   - exposure linked to pathogenesis

6. Consideration of alternative explanations
   - the extent to which other explanations have been considered.

Nature of Evidence

7. Cessation of exposure (Dynamics)
   - removal of exposure – reduces risk

8. Specificity
   - specific exposure is associated with only one disease

9. Experimental evidence

Practice 3

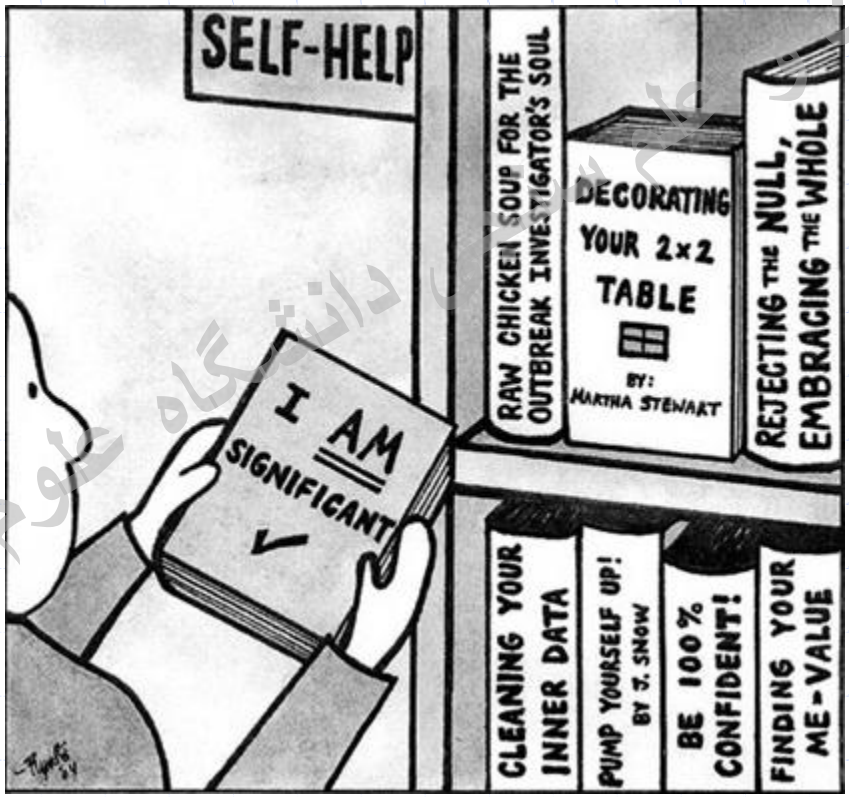
Statistical methods

Why does good evidence from research fail to get into practice??

- 75% cannot understand the statistics

- 70% cannot critically appraise a research paper

Dunn, Virginia, et al. "Using research for practice: a UK experience of the BARRIERS Scale." *Journal of Advanced Nursing* 26.6 (1997): 1203-1210.



Why is statistics necessary?

- 58% of the population had GERD
- Mean age of the respondents was 25±8
- 25% of women and 50% of men lied about their age.
- Doctors live longer than normal people

Why is statistics necessary?

- Descriptive statistics
  - 58% of the population had GERD
  - Mean age of the respondents was 25±8
- Analytical statistics
  - 25% of women and 50% of men lied about their age
  - Doctors live longer than normal people.

Descriptive statistics

- Depend on the type of the variables use:
  - number and percentage
  - the mean and its standard deviation
  - the median and its inter-quartile range

Inferential statistics: exploring associations and differences

Differences

- Continuous variables (blood pressure, age): 109 vs. 140
- Categorical variables (proportion of blind people): 10% vs. 2%

Measures of Association

Relative Risk (RR)

Odds Ratio (OR)

Attributable Risk

Treatment	dead	alive	Total
Medical	404	921	1325
CABG	350	974	1324

Measures of Association

- Linear Correlation
  - r
- Regression
  - Univariate
  - Multiple Regression
  - Logistic Regression
  - Cox Proportional Hazard Model
- Do they mean causation?

- What is the appropriate test?
  - Scales
    - Nominal
    - Ordinal
    - Interval
    - Ratio

Statistical Tests

	Type of Variable		
Goal	**Quantitative**	**Ordinal**	**categorical**
Describe one group	Mean, SD	Median, interquartile range	Proportion
Compare one group to a hypothetical value	One-sample *t test*	Wilcoxon test	Chi-square or Binomial test **
Compare two unpaired *groups*	Unpaired *t test*	Mann-Whitney test	Fisher's test (chi-square for large samples)
Compare two paired groups	Paired *t test*	Wilcoxon test	McNemar's test

Statistical Tests

Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q**
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients**
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression**	Simple logistic regression*
Predict value from several measured or binomial variables	Multiple linear regression* or Multiple nonlinear regression**		Multiple logistic regression*

Practice 4

بررسی قسمتهای مختلف مقاله

1.Check the Title

- Read the title and check that you understand its meaning. Sometimes titles are inaccurate and do not reflect the content of the paper which follows.

- For example, one title indicating the use of a drug in the treatment of hypertension, prefaced a paper which merely described a short haemodynamic study.

1.Check the Title

- Watch for cryptic titles. Sometimes a useful paper may be hidden behind an indifferent title.

- Never rely on the title alone to accept or reject a paper for more detailed reading.

2.Who are the Authors?

- Range of expertise: professional backgrounds with address
- Research center?
- Principle researcher
- Number of authors
- Have any of the authors obvious connections with the drug industry?

3.Read the abstract

- This is a synopsis of the paper, which should give the objective of the study, the methods used, the results obtained and the conclusions reached.

3.Read the abstract

Beware of the following warning signs:

1. Confusion and possible contradictory statements - a good abstract should be crystal clear.

2. **Overuse** of statistical terms (especially p values).

3. Disparity between the number of subjects mentioned in the summary and the number in the paper

4.Check the Introduction

- Check that a brief review of available background literature is provided and that the question being asked in the study follows logically from the available evidence.

Introduction

- General, concise description of problem
  - background to the work
  - previous research
- Where that work is deficient
  - how your research will be better
- State the hypothesis
- About 3 to 4 paragraphs

Methods

- Study design
- Participants
- Ethical approval
- Sample size
- Questionnaires
- Interventions
- Clinical assessments
- Statistical methods

6. Results

What was found?

- Should be logical – simple → complex

Cheat on statistical tests

- If baseline differences between the groups favour the intervention group, remember not **to** adjust for them

- Ignore all withdrawals (drop outs) and non-responders, so the analysis only concerns subjects who fully complied with treatment

Cheat on statistical tests

- Always assume that you can plot one set of data against another and calculate an "*r* value" (Pearson correlation coefficient), and assume that a "significant" *r* value proves causation
- If outliers (points which lie a long way from the others on your graph) are messing up your calculations, just rub them out. But if outliers are helping your case, even if they seem **to** be spurious results, leave them in
- If the confidence intervals of your result overlap zero difference between the groups, leave them out of your report.

- If the difference between two groups becomes significant four and a half months into a six month trial, stop the trial and start writing up. Alternatively, if at six months the results are "nearly significant," extend the trial for another three weeks

- If your results prove uninteresting, ask the computer to go back and see if any particular subgroups behaved differently. You might find that your intervention worked after all in Chinese women aged 52-61

Does the y-axis start at zero?

- The y-axis should always begin at zero. If this is not so, someone is trying to make you believe that one of the groups has reached the lowest rate or number possible when this is not the case.

7. Discussion

- Check that the progress in argument to the conclusion is logical and also that any doubts or inconsistencies which have been raised in your mind by earlier parts of the paper, are dealt with.
- Are limitations mentioned?

8.Bibliography

- If you find statements in the paper which you consider to be important check that a reference is provided.
- Be suspicious if no reference is given, or if the references which are provided are dated, or predominantly in obscure journals.

9. Acknowledgment

- Who? (and what)?
- Source of funding? (conflict of interests)

Recommended Reading

- Trisha Greenhalgh : How to read a paper; the basis of evidence based medicine

- Gordon Guyatt, Drummond Rennie. Users' Guides To The Medical Literature, A Manual for Evidence-Based Clinical Practice

CHECK-LISTS AND TOOLS

What is critical appraisal?

- **Critical appraisal** is the use of explicit, transparent methods to assess the data in published research, applying the rules of evidence to factors such as internal validity, adherence to reporting standards, conclusions and generalizability

Critical appraisal is **not**:	Critical appraisal is:
× Negative dismissal of any piece of research	✓ Balanced assessment of benefits and strengths of research against its flaws and weaknesses
× Assessment of results alone	✓ Assessment of research process and results
× Based entirely on detailed statistical analysis	✓ Consideration of quantitative and qualitative aspects of research
× To be undertaken by expert researchers/statisticians only	✓ To be undertaken by all health professionals as part of their work

Critical Appraisal: Three preliminary questions

- **Why** was the study done and what hypothesis was being tested?
- **What** type of study was done?
- **Was** the study **design appropriate**?

Key Steps To Effective Critical Appraisal

1. What are the results?
2. Are the Results valid?
3. How will these results help me/my colleagues do their job?

Critical Appraisal Tools

- Why do we need them?
- Where we can find them?

EQUATOR network

Enhancing the QUAlity and Transparency Of health Research

http://www.equator-network.org/

www.equator-network.org

equator network

Enhancing the QUAlity and Transparency Of health Research

EQUATOR resources in Portuguese | Spanish

Home | Library | Toolkits | Courses & events | News | Blog | About us | Contact

Essential resources for writing and publishing health research

Library for health research reporting

The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to research reporting.

- Search for reporting guidelines
- Not sure which reporting guideline to use?
- Reporting guidelines under development
- Visit the library for more resources

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions	Other
Observational studies	STROBE	Extensions	Other
Systematic reviews	PRISMA	Extensions	Other
Case reports	CARE		Other
Qualitative research	SRQR	COREQ	Other
Diagnostic / prognostic studies	STARD	TRIPOD	Other
Quality improvement studies	SQUIRE		Other
Economic evaluations	CHEERS		Other
Animal pre-clinical studies	ARRIVE		Other
Study protocols	SPIRIT	PRISMA-P	Other

See all 319 reporting guidelines

Reporting guidelines highlighted in a new report on reproducibility and reliability of

History

The EQUATOR programme grew out of the work of CONSORT and other guideline development groups. The **project began in March 2006**. Initially funded for one year by the UK NHS National Knowledge Service, the project had three major objectives: to map the current status of all activities aimed at preparing and disseminating guidelines on reporting health research studies, identify key individuals working in the area, and establish relationships with potential key stakeholders.

The EQUATOR Network held its **first international working meeting** in Oxford in May-June 2006, attended by 27 participants from 10 countries. The participants included representatives of reporting guideline development groups, journal editors, peer reviewers, medical writers and funders. The objective of the meeting was to exchange experience in developing, using and implementing reporting guidelines and outline priorities for the future EQUATOR Network activities.

Prior to the first EQUATOR meeting we searched literature to identify published reporting guidelines and surveyed authors to examine how the guidelines were developed and to identify problems encountered during the development (see Simera et al, PLoS Med 2008).

The survey results and meeting discussions helped us to prioritise main activities that were necessary for a successful

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Case reports	CARE	
Qualitative research	SRQR	COREQ
Diagnostic / prognostic studies	STARD	TRIPOD
Quality improvement studies	SQUIRE	
Economic evaluations	CHEERS	
Animal pre-clinical studies	ARRIVE	
Study protocols	SPIRIT	PRISMA-P

www.equator-network.org/about-us/

equator network

Enhancing the QUAlity and Transparency Of health Research

EQUATOR resources in Portuguese | Spanish

Home | Library | Toolkits | Courses & events | News | Blog | About us | Contact

Home > About us

About us

The EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network is an international initiative that seeks to improve the reliability and value of published health research literature by promoting transparent and accurate reporting and wider use of robust reporting guidelines.

It is the first coordinated attempt to tackle the problems of inadequate reporting systematically and on a global scale; it advances the work done by individual groups over the last 15 years.

- EQUATOR Network: what we do and how we are organised
- History of EQUATOR
- UK EQUATOR Centre
- Canadian EQUATOR Centre
- French EQUATOR Centre

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Case reports	CARE	
Qualitative research	SRQR	COREQ
Diagnostic / prognostic studies	STARD	TRIPOD
Quality improvement studies	SQUIRE	
Economic evaluations	CHEERS	
Animal pre-clinical studies	ARRIVE	
Study protocols	SPIRIT	PRISMA-P

- **EQUATOR Network: what we do and how we are organised**

The EQUATOR Network is an 'umbrella' organisation that brings together researchers, medical journal editors, peer reviewers, developers of reporting guidelines, research funding bodies and other collaborators with mutual interest in improving the quality of research publications and of research itself.

In 2014 we launched the first three centres that will substantially contribute to expanding the EQUATOR activities: the UK EQUATOR Centre (also the EQUATOR Network's head office), French EQUATOR Centre and Canadian EQUATOR Centre.

The new centres will focus on national activities aimed at raising awareness and supporting adoption of good research reporting practices.

EQUATOR's mission and goals

- The EQUATOR mission is to achieve accurate, complete, and transparent reporting of all health research studies to support research reproducibility and usefulness.
- Our work increases the value of health research and helps to minimize avoidable waste of financial and human investments in health research projects.

- To achieve its mission the EQUATOR Network has the following major goals:

Maintain and further develop a comprehensive collection of online resources providing up-to-date information, tools and other materials related to health research reporting (Library for health research reporting)

EQUATOR's mission and goals

- Actively promote the use of reporting guidelines and good research reporting practices through an education and training programme
- Assist in the development, dissemination and implementation of robust reporting guidelines
- Support journals, universities and other organisations in implementing reporting guidelines through development of tools, strategies, education and other activities
- Undertake research projects enhancing the value of health-related research
- Set up a global network of local EQUATOR centres to facilitate the improvement of health research reporting on a worldwide scale

Steering Group members:

- **Doug Altman**, Director, Centre for Statistics in Medicine, Oxford, UK (Chair); Director UK EQUATOR Centre
- **Trish Groves**, Head of Research, BMJ & Editor-in-chief, BMJ Open
- **Ana Marušić**, Professor and Chair, Department of Research in Biomedicine and Health, University of Split, Croatia
- **David Moher**, Senior Scientist, Ottawa Health Research Institute, Ottawa, Canada; Director, Canadian EQUATOR Centre
- **Philippe Ravaud**, Director, Centre of Epidemiology at the Hotel-Dieu (Paris); Director, French EQUATOR Centre
- **Iveta Simera**, Deputy Director UK EQUATOR Centre / Programme Manager, EQUATOR Network, Centre for Statistics in Medicine, Oxford, UK

Critical Appraisal Skills Programme (CASP)

http://www.casp-uk.net/

Critical Appraisal Skills Programme (CASP)

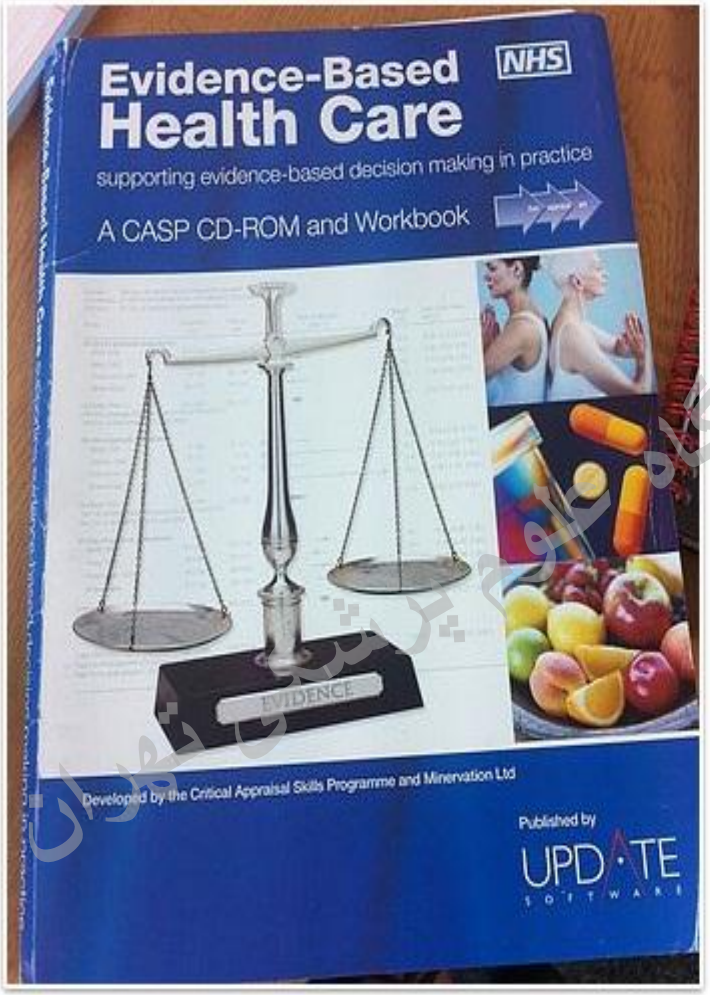
Making sense of evidence

HOME | CRITICAL APPRAISAL | WORKSHOPS | CASP TOOLS & CHECKLISTS | ABOUT CASP | More

CASP offers critical appraisal skills training, workshops and tools. These help you read and check health research for trustworthiness, results & relevance.

Sign up here to find out about upcoming CASP workshops and events

First Name:



HISTORY

CASP was initiated under Sir Muir Gray when he was Director of Research & Development at Oxford Regional Health Authority in 1993.

It was in response to the need for developing skills in health care staff to meet the challenge of Evidence Based Medicine.

The workshop format was developed by trial and error with willing guinea pigs! The core CASP checklists (randomised controlled trial & systematic review) were based on JAMA 'Users' guides to the medical literature 1994 (adapted from Guyatt GH, Sackett DL, and Cook DJ), and piloted with health care practitioners.

For each new checklist (e.g qualitative) a group of experts were assembled to develop and pilot the checklist and the workshop format with which it would be used.

Over the years overall adjustments have been made to the format, but a recent survey of checklist users reiterated that the basic format continues to be useful and appropriate.

WHO IS CASP FOR?

CASP is for anyone that wants to use research evidence in their professional practice, professional and personal decision making, and policy & guidelines development.

Scenarios:

- Your clinical department wants to improve the organisation of the outpatient clinic, and you have found a systematic review of relevance and a recent patient survey report. The clinical management team meet in 3 weeks to discuss potential changes. Two members of the group critically appraise the review and read the survey report independently, and discuss their findings ahead of the meeting. At the beginning of the team meeting they report back on the review and its findings, these underpin the next stage of discussion about reconfiguring the clinic that includes the results of a patient survey and the views & experiences of the management team.

- Your elderly parent needs a hip replacement, he is frail and anxious about having surgery. Prior to an appointment with the orthopaedic consultant you find a

- You are a new member of the public health team at 'Anywhere Council'. The council are

- You are a new member of the public health team at 'Anywhere Council'. The council are discussing whether to continue a subsidised exercise programme for overweight teenagers living in 'Anywhere'. Ahead of the council meeting you find a systematic review and often cited qualitative paper that is relevant to the discussion, you appraise these papers and prepare a short presentation about its findings and recommendations. You request a slot on the agenda entitled "What is the evidence of effectiveness of community exercise interventions for overweight teenagers."

Scenarios:

- Your clinical department wants to improve the organisation of the outpatient clinic, and you have found a systematic review of relevance and a recent patient survey report. The clinical management team meet in 3 weeks to discuss potential changes. Two members of the group critically appraise the review and read the survey report independently, and discuss their findings ahead of the meeting. At the beginning of the team meeting they report back on the review and its findings, these underpin the next stage of discussion about reconfiguring the clinic that includes the results of a patient survey and the views & experiences of the management team.

- Your elderly parent needs a hip replacement, he is frail and anxious about having surgery. Prior to an appointment with the orthopaedic consultant you find a systematic review about the effectiveness of hip replacement surgery for osteoarthritis. Using the CASP checklist you appraise this review, especially the outcomes that are being measured, and take this along to the appointment to discuss further.

CASP ...

- Systematic Reviews
- Randomized Controlled Trials (RCTs)
- Qualitative Research
- Economic Evaluation Studies
- Cohort Studies
- Case Control Studies
- Diagnostic Test Studies

Three questions

- **Valid?**

  Is the methodology appropriate to answer the question.

  Is it carried out in a sound way, eliminating bias and confounding?

- **Result?**
- What are the result?
- **Applicable?**

  Will the results help locally?

International Centre for Allied Health Evidence

- Appraising randomized controlled trials
- Appraising non-randomized controlled trials
- Appraising other forms of quantitative research
- Appraising case studies
- Appraising qualitative research
- Appraising mixed methods research
- Appraising systematic reviews
- Appraising meta-analyses
- Appraising clinical guidelines
- Appraising outcome measures
- Assessing treatment choices

www.unisa.edu.au/Research/Sansom-Institute-for-Health-Research/Research/Allied-Health-Evidence/Resources/

CELEBRATING 25 YEARS

University of South Australia

STUDY RESEARCH PARTNER NEWS & EVENTS LOGIN

Home > Research > Sansom Institute for Health Research > International Centre for Allied Health Evidence > Resources

International Centre for Allied Health Evidence

- Allied and Scientific Health News
- Resources
  - Critical Appraisal Tools
  - Glossary of terms
  - Guideline Clearinghouse
  - iCAHE Journal Clubs
  - iCAHE Journal Club Critical Appraisals
  - iCAHE Masterclass
  - iCAHE Outcome Calculators
  - iCAHE textbooks
  - Useful websites
  - iCAHE's Learning Hub
- Quality Care
- UniSA Staff Members of iCAHE
- iCAHE Research Areas

Resources

This website offers freely-available access to a range of resources developed by iCAHE over time. These resources have come from projects that iCAHE researchers have conducted and from requests for information from clients, collaborators and associates. This page contains links to an ongoing and constantly evolving collection of resources designed to promote the continual improvement of quality and safety of allied health care, that is at the heart of all iCAHE activities. The core theme running throughout these resources is ease of access, to aid implementation of evidence into practice. We anticipate that this collection of resources will meet the needs of clinicians, researchers, educators, students and consumers of health care for free and easy access to relevant information.

Who are you?

Clinicians

Are you a health care practitioner/ provider

Researchers

Are you a researcher who is interested in

Educators and Students

Are you an

Consumers

Are you a consumer of health care who is

AGREE

- **Appraisal of Guidelines Research and Evaluation**
- The AGREE Instrument for the assessment of clinical practice guidelines is available on-line in several languages

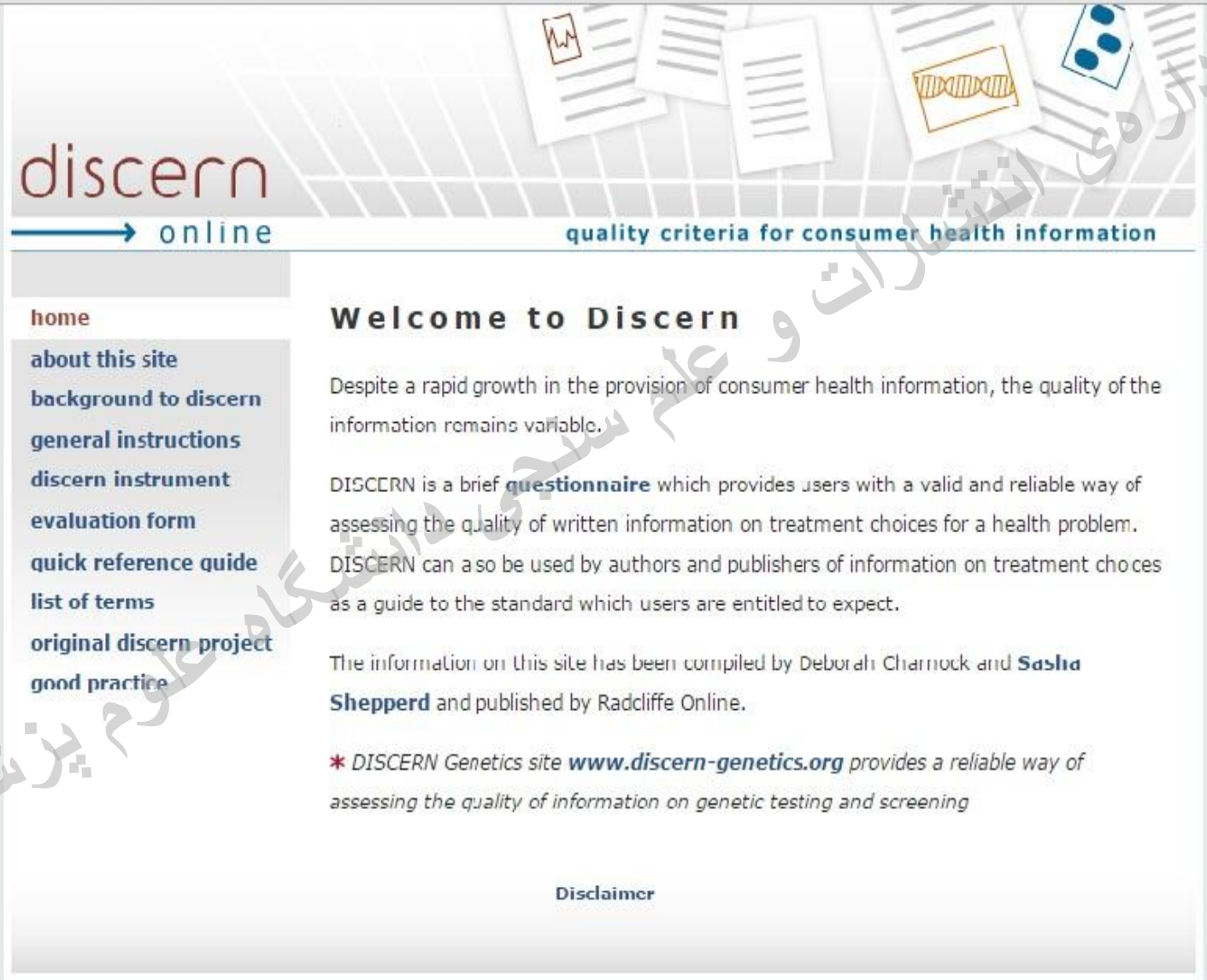
  http://www.agreecollaboration.org

DISCERN

Quality criteria for consumer health information on treatment choices

- DISCERN is a brief questionnaire which provides users with a valid and reliable way of assessing the quality of written information on **treatment choices** for a health problem.
- DISCERN can also be used by authors and publishers of information on treatment choices as a guide to the standard which users are entitled to expect.
- New DISCERN Genetics site provides a reliable way of assessing the quality of information on genetic testing and screening.

www.discern.org.uk



STROBE Statement

- **STrengthening the Reporting** of **OBservational studies in Epidemiology**
- STROBE stands for an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of observational studies, with the common aim of STrengthening the Reporting of OBservational studies in Epidemiology.

- **www.strobe-statement.org**

Do they work?

- Katrak et al. **systematic review of the content of critical appraisal tools.** *BMC Medical Research Methodology 2004*

- Few critical appraisal tools had documented evidence of validity of their items, or reliability of use.

Appraisal Tools for

OBSERVATIONAL STUDIES

Types of Observational studies

- Cohort
- Case-control
- Cross-sectional
- Ecologic
- Case series
- Case report

STROBE Statement

- STrengthening the Reporting of OBservational studies in Epidemiology
- Many journals refer to the STROBE Statement in their Instructions for Authors.
- Provides recommendation for each section (22 items)

Available STROBE check-lists

- STROBE checklist for **cohort, case-control, and cross-sectional studies** (combined)
- Checklist for **cohort studies**
- Checklist for **case-control studies**
- Checklist for **cross-sectional studies**

Title and abstract

(a) Indicate the study's design with a commonly used term in the title or the abstract

(b) Provide in the abstract an informative and balanced summary of what was done and what was found

Introduction

- **Background/rationale:**
  - Explain the scientific background and rationale for the investigation being reported
- **Objectives:**
  - State specific objectives, including any pre-specified hypotheses

Methods

- **Study design**
  - Present key elements of study design early in the paper
- **Setting**
  - Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection

Methods: participants

- **Cohort study**—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up
- **Case-control study**—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls
- **Cross-sectional study**—Give the eligibility criteria, and the sources and methods of selection of participants

Methods: matched studies

- Cohort study—For matched studies, give matching criteria and number of exposed and unexposed

- Case-control study—For matched studies, give matching criteria and the number of controls per case

Methods: Variables

- Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable

- **Quantitative variables**
  - Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why

Methods: Data sources/ measurement

- For each variable of interest, give sources of data and details of methods of assessment (measurement).

- Describe comparability of assessment methods if there is more than one group

Method: Bias & Study size

- Describe any efforts to address potential sources of bias

- Explain how the study size was arrived at

Method: Statistical methods

- Describe all statistical methods, including those used to control for confounding
- Describe any methods used to examine subgroups and interactions
- Explain how missing data were addressed
- **Cohort study**—If applicable, explain how loss to follow-up was addressed
- **Case-control study**—If applicable, explain how matching of cases and controls was addressed
- **Cross-sectional study**—If applicable, describe analytical methods taking account of sampling strategy
- Describe any sensitivity analyses

Results: Participants

- Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed
- Give reasons for non-participation at each stage
- Consider use of a flow diagram

Results: Descriptive data

- characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders
- number of participants with missing data for each variable of interest
- *Cohort study—Summarise follow-up time (eg, average and total amount)*

Results: Outcome data

- Cohort study—Report numbers of outcome events or summary measures over time
- Case-control study—Report numbers in each exposure category, or summary measures of exposure
- Cross-sectional study—Report numbers of outcome events or summary measures

Main results and Other analyses

- unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
- Report category boundaries when continuous variables were categorized
- If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
- Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

Discussion

- **Key results**: Summarize key results with reference to study objectives
- **Limitations**: Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
- **Interpretation**: Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
- **Generalisability**: Discuss the generalisability (external validity) of the study results

Other information

- the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

CASP: Cohort study

CRITICAL APPRAISAL SKILLS PROGRAMME
making sense of evidence

12 questions to help you make sense of a cohort study

Public Health Resource Unit, Oxford

General comments

- Three broad issues need to be considered when appraising a cohort study:

  - *Are the results of the study valid?*
  - *What are the results?*
  - *Will the results help locally?*

screening questions

- The first two questions are screening questions and can be answered quickly. If the answer to those two is "yes", it is worth proceeding with the remaining questions.

1. **Did the study address a clearly focused issue?**

2. **Did the authors use an appropriate method to answer their question?**

(A) Are the results of the study valid?

1,2. screening questions

3. Was the cohort recruited in an acceptable way?

4. Was the exposure accurately measured to minimize bias?

5. Was the outcome accurately measured to minimize bias?

6. Have the authors identified all important confounding factors? Have they taken account of the confounding factors in the design and/or analysis?

7. Was the follow up of subjects complete enough? Was the follow up of subjects long enough?

What are the results?

8. What are the results of this study?

9. How precise are the results? How precise is the estimate of the risk?

10. Do you believe the results?

Will the results help me locally?

11. Can the results be applied to the local population?

12. Do the results of this study fit with other available evidence?

Practice 5

Appraisal Tools for

RANDOMIZED CONTROLLED TRIALS

CONSORT

- Consolidated Standards of Reporting Trials
- 25 items

HISTORY OF CONSORT

- CONSORT (Consolidated Standards of Reporting Trials) statement (In the mid 1990s)
- The revised CONSORT statement (1999, 2000)
- CONSORT 2010

The CONSORT statement comprises:

a 25-item checklist **pertain to the content of**

the Title,
Abstract,
Introduction,
Methods,
Results,
discussion

Other information

a flow diagram **depicts information from 4 stages of a trial**

enrollment,
intervention allocation,
follow-up,
analysis

Title and abstract

- How participants were allocated to interventions (e.g., “random allocation,” “randomized,”or “randomly assigned”).

Introduction: Background

- Scientific background and explanation of rationale.

Method:

- **Participants**: Eligibility criteria for participants and the settings and locations where the data were collected.

- **Interventions:** Precise details of the interventions intended for each group and how and when they were actually administered.

- **Objectives:** Specific objectives and hypotheses.

Method:

- **Outcomes:** Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).

- **Sample size:** How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.

Method: Randomization

- **Sequence generation:** Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking, stratification).
- **Allocation concealment:** Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.
- **Implementation:** Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.

Method:

- **Blinding** (masking): Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.

- **Statistical methods:** Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses.

Results

- **Participant flow:** Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.

- **Recruitment:** Dates defining the periods of recruitment and follow-up.

- **Baseline data:** Baseline demographic and clinical characteristics of each group.

Results

- **Numbers analyzed:** Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat." State the results in absolute numbers when feasible (e.g., 10 of 20, not 50%).

- **Outcomes and estimation:** For each primary and secondary outcome, a summary of results for each group and the estimated effect size and its precision (e.g., 95% confidence interval).

Results

- **Ancillary analyses:** Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.

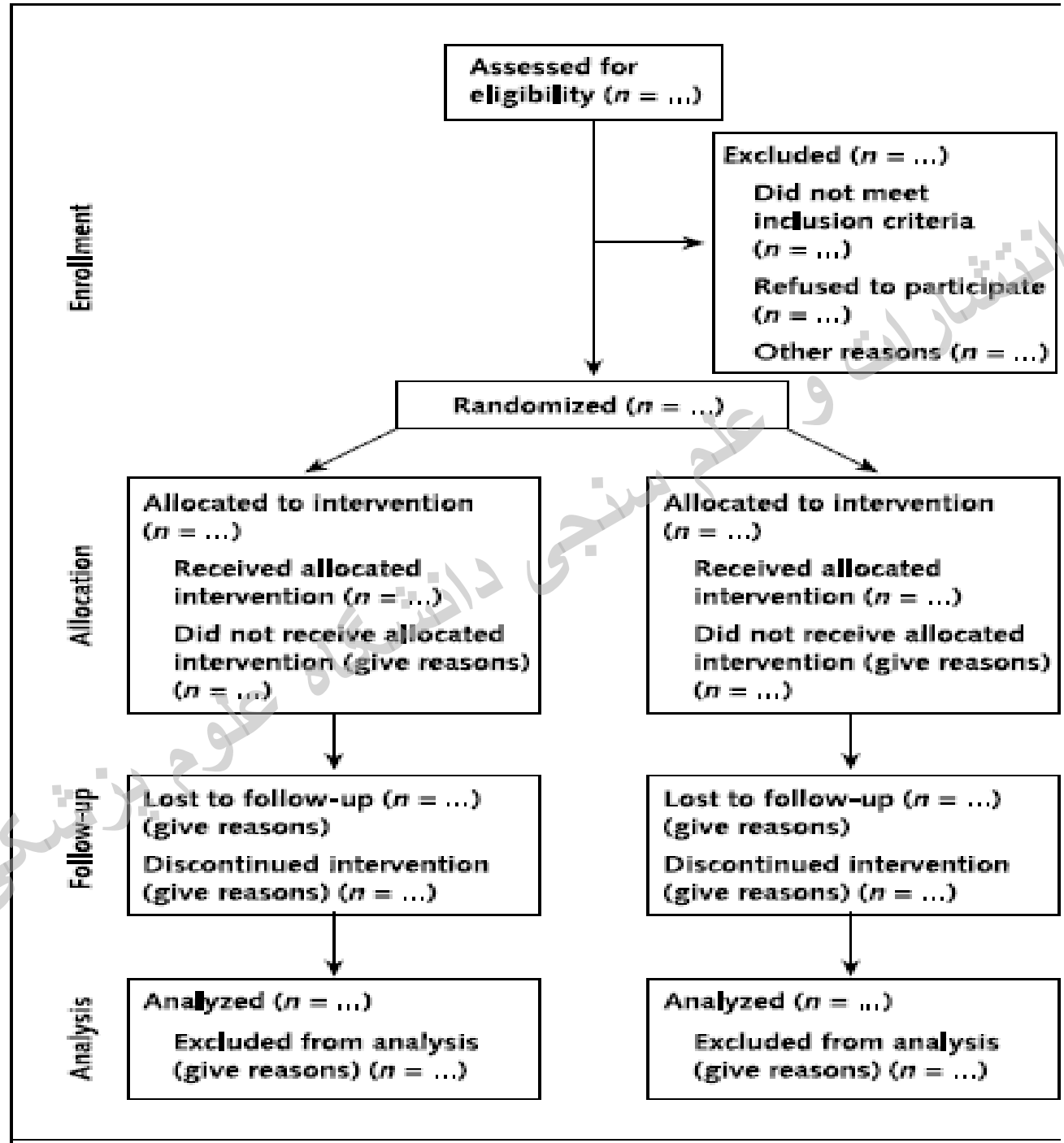
- **Adverse events:** All important adverse events or side effects in each intervention group

Discussion

- **Interpretation:** Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes.

- **Generalizability:** Generalizability (external validity) of the trial findings.

- **Overall evidence:** General interpretation of the results in the context of current evidence.



CASP

11 questions to help you make sense of a trial

How to use this appraisal tool

Three broad issues need to be considered when appraising the report of a randomised controlled trial:

- **Are the results of the trial valid?** **(Section A)**
- **What are the results?** **(Section B)**
- **Will the results help locally?** **(Section C)**

(A) Are the results of the trial valid?

Screening questions:

1.Did the trial address a clearly focused issue?

2. Was the assignment of patients to treatments randomised?

- Consider:
- How was this carried out, some methods
- may produce broken allocation concealment
- Was the allocation concealed from researchers?

Screening Questions

1. Did the trial address a clearly focused issue? ☐ Yes ☐ Can't tell ☐ No

Consider: An issue can be 'focused' in terms of

- The population studied
- The intervention given
- The comparator given
- The outcomes considered

---

2. Was the assignment of patients to treatments randomised? ☐ Yes ☐ Can't tell ☐ No

Consider:

- How was this carried out, some methods may produce broken allocation concealment
- Was the allocation concealed from researchers?

Is it worth continuing?

Detailed questions:

- **Are the results of the trial valid?**

1 ,2. Screening Questions

3. Were patients, health workers and study personnel blinded?
4. Were the groups similar at the start of the trial?
5. Aside from the experimental intervention, were the groups treated equally?
6. Were all of the patients who entered the trial properly accounted for at its conclusion?

Detailed questions:

B: what are the results:

7. How large was the treatment effect?
8. How precise was the estimate of the treatment effect?

C:Will the results help locally?

9. Can the results be applied in your context? (or to the local population?)
10. Were all clinically important outcomes considered?
11. Are the benefits worth the harms and costs?

Practice 6

Appraisal Tools for

DIAGNOSTIC TESTS

Kamran Yazdani, MD MPH PhD
Assistant Professor in Epidemiology
Tehran University of Medical Sciences

Diagnostic tests

- When looking at a paper about a diagnostic test we ask ourselves three questions.

Diagnostic tests

- Is this test useful?

Diagnostic tests

- Is this test useful?
- Is it reliable?

Diagnostic tests

- Is this test useful?
- Is it reliable?
- Is it valid?

Is this test useful?

- The test should have been researched in a study population **relevant** to the individual or population in whom it is to be used.

Reliability

- Reliability refers to the **repeatability** or reproducibility of a test.
- It can be assessed by repeating the test using the same or different observers.

Validity

- Relates to whether the test measures what it purports to measure. Is the result true?

- It can be assessed by comparing the test results with a Gold Standard.

Validity

- For example if you measure blood pressure in an obese patient and use a cuff that is too small you are likely to get a falsely high reading. The reading maybe reliable (you get the same blood pressure if you do it again) but it lacks validity.

Gold standard

- The gold standard is the test or battery of tests that will most accurately diagnose a particular disease or condition.
  - The OGTT for diabetes
  - Fluoroscein angiography for diabetic retinopathy (too expensive or invasive)
  - The Jones criteria for rheumatic fever (a battery of tests or symptoms)

Which one is BETTER?

What is the accuracy?

Indices for the assessment of Validity and Reliability

The type of variable?

Table 8–3 Summary of Indices or Graphic Approaches Most Frequently Used for the Assessment of Validity and Reliability

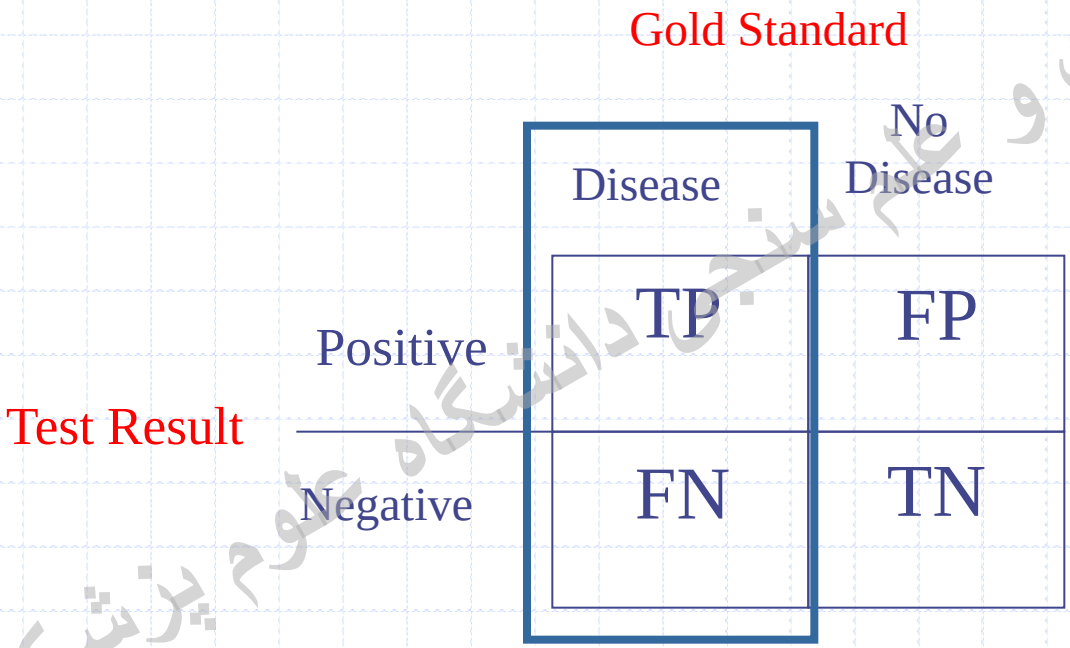
Type of Variable	*Index or Technique*	*Mostly Used to Assess . . .*	
		Validity	*Reliability*
Categorical	Sensitivity/specificity	++	
	Percent agreement	+	++
	Percent positive agreement	+	++
	Kappa statistic	+	++
Continuous	Scatter plot (correlation graph)	+	++
	Linear correlation coefficient (Pearson)	+	+
	Ordinal correlation coefficient (Spearman)	+	+
	Intraclass correlation coefficient	+	++
	Coefficient of variation		++
	Bland-Altman plot	++	++

Note: ++, the index is indicated and used to measure the magnitude of validity or reliability; +, although the index is used to measure the magnitude of either validity or reliability, its indication is somewhat questionable.

Ability of a test to accurately diagnose diseased and healthy individuals

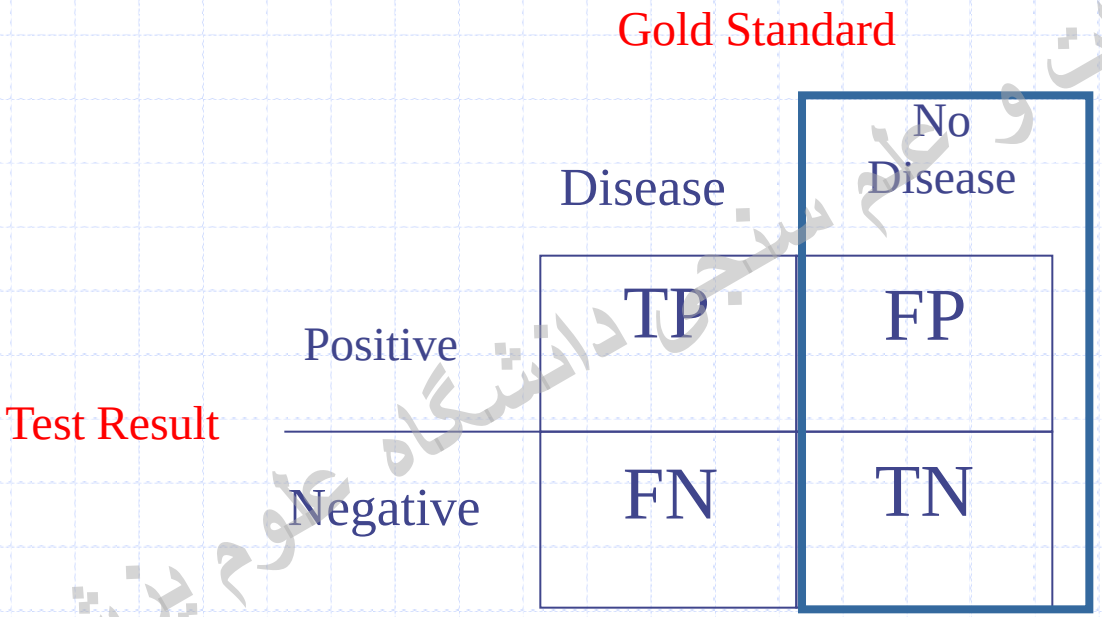
- Sensitivity
- Specificity

Sensitivity



Sensitivity: The capacity of the test to correctly identify **diseased** individuals in a population; "TRUE POSITIVES".

Specificity



Specificity: The capacity of the test to correctly exclude individuals who are **free of the disease;** "TRUE NEGATIVES".

Sensitivity and Specificity

Test Result	Gold Standard: Disease	Gold Standard: No Disease
Positive	TP	FP
Negative	FN	TN

Example

Test Result	Gold Standard: Disease	Gold Standard: No Disease	
Positive	75	20	95
Negative	25	180	205
	100	200	300

Sensitivity = ???

Specificity = ???

Example

Test Result	Gold Standard: Disease	Gold Standard: No Disease	
Positive	75	20	95
Negative	25	180	205
	100	200	300

Sensitivity = 75/100 = 75%

Specificity = 180/200 = 90%

Accuracy of the test

Test Result	Gold Standard: Disease	Gold Standard: No Disease	
Positive	a	b	a+b
Negative	c	d	c+d
	a+c	b+d	300

What do we do in clinic?

- Positive & Negative Predictive Values
  - Using Sens, Spec, and Prevalence to calculate
- Likelihood Ratio
  - Using Sens, and Spec to calculate
  - Using pre-test prob. to predict post-test prob.

Positive Predictive Value

Test Result	Gold Standard: Disease	Gold Standard: No Disease
Positive	TP	FP
Negative	FN	TN

PPV: The probability of the disease being present, among those with positive diagnostic test results

Negative Predictive Value

Test Result	Gold Standard: Disease	Gold Standard: No Disease
Positive	TP	FP
Negative	FN	TN

NPV: The probability that the disease was absent, among those whose diagnostic test results were negative

The effect of Sens, Spec, and P on PPV and NPV

		PPV			NPV		
		Prevalence					
Sensitivity	Specificity	90%	50%	10%	90%	50%	10%
70%	60%	94%	64%	16%	18%	67%	95%
70%	90%	98.4%	88%	44%	25%	75%	96%
80%	90%	98.6%	89%	47%	33%	82%	98%
90%	90%	98.7%	90%	50%	50%	90%	99%
100%	5%	2%	51%	10%	100%	100%	100%
5%	100%	100%	100%	100%	98%	51%	90%

Are there some predictors other than the prevalence?

Likelihood ratio

$$\text{LR Positive} = \frac{\text{Likelihood of (+) test in diseased persons}}{\text{Likelihood of (+) test in healthy persons}}$$

$$\text{LR Positive} = \frac{\text{Sensitivity}}{\text{1 - Specificity}}$$

$$\text{LR Negative} = \frac{\text{Likelihood of (-) test in diseased persons}}{\text{Likelihood of (-) test in healthy persons}}$$

$$\text{LR Negative} = \frac{\text{1 - Sensitivity}}{\text{Specificity}}$$

Likelihood ratio

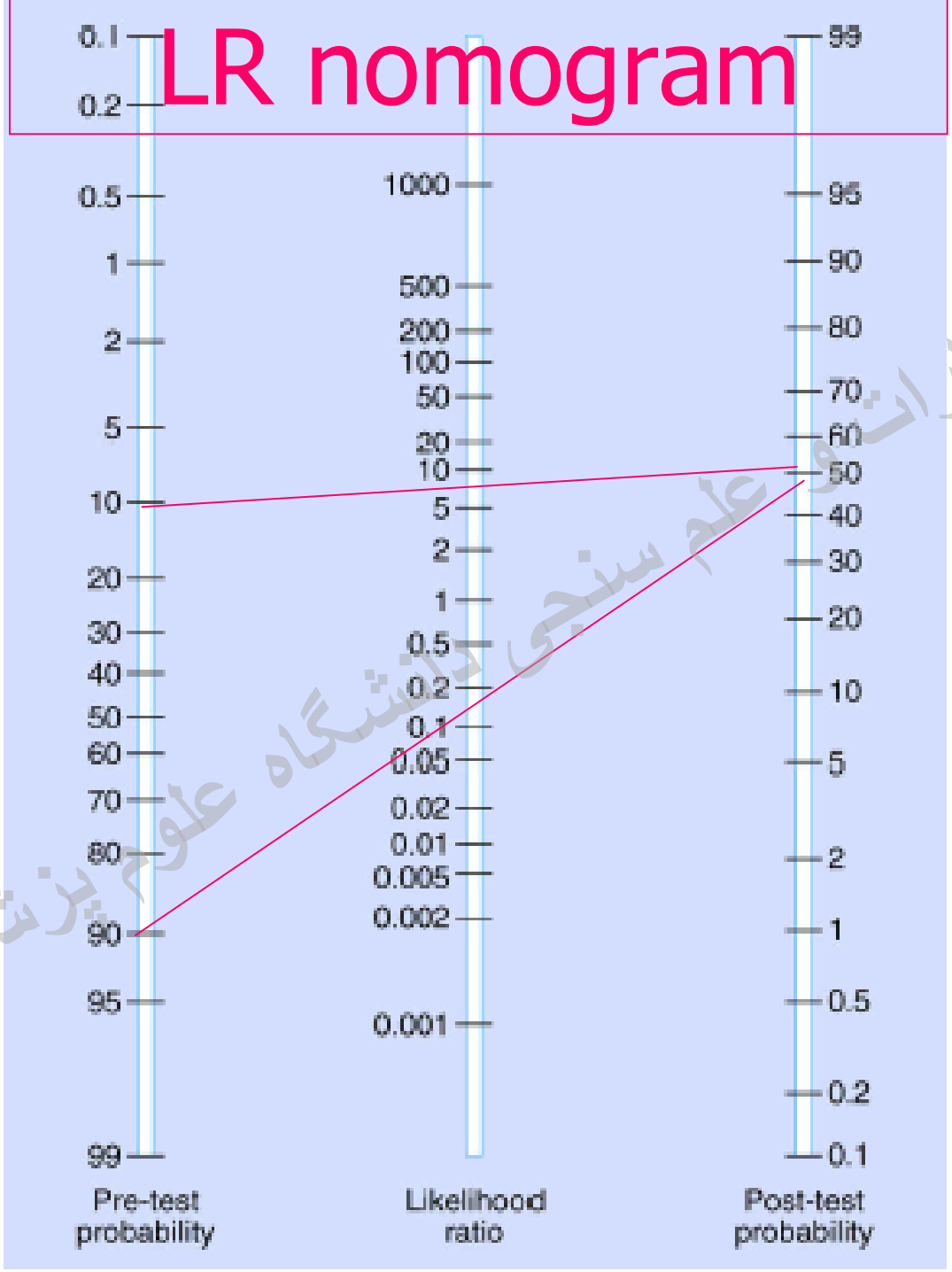
Sensitivity = 90%

Specificity = 90%

$$\text{LR Positive} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} = \frac{0.90}{1 - 0.90} = 9$$

$$\text{LR Negative} = \frac{1 - \text{Sensitivity}}{\text{Specificity}} = \frac{1 - 0.90}{0.90} = 1/9$$

LR nomogram



5000 pregnant women underwent a test for blood glucose at 24 weeks, following a glucose load. 243 women were found to have a blood glucose greater than 6.8 mmol/L and were referred for an OGTT. 186 were found to have gestational diabetes. Four women who initially had tested negative were diagnosed as having diabetes later in their pregnancy.

	Diabetes	No diabetes	Total
Positive	186	57	243
Negative	4	4753	4757
Total	190	4810	5000

Example

- Prevalence
- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value
- Likelihood ratio + test
- Likelihood ratio - test
- Accuracy

Prevalence	190/5000
Sensitivity	186/190
Specificity	4753/4810
Positive predictive value	186/243
Negative predictive value	4753/4757
Likelihood ratio + test	(186/190)/(57/4810)
Likelihood ratio - test	(4/190)/(4753/4810)
Accuracy	(186+4753)/5000

Prevalence	3.8%
Sensitivity	97.9%
Specificity	98.8%
Positive predictive value	76.5%
Negative predictive value	99.9%
Likelihood ratio + test	82.6
Likelihood ratio - test	.02
Accuracy	98.8%

Continuous Measurements

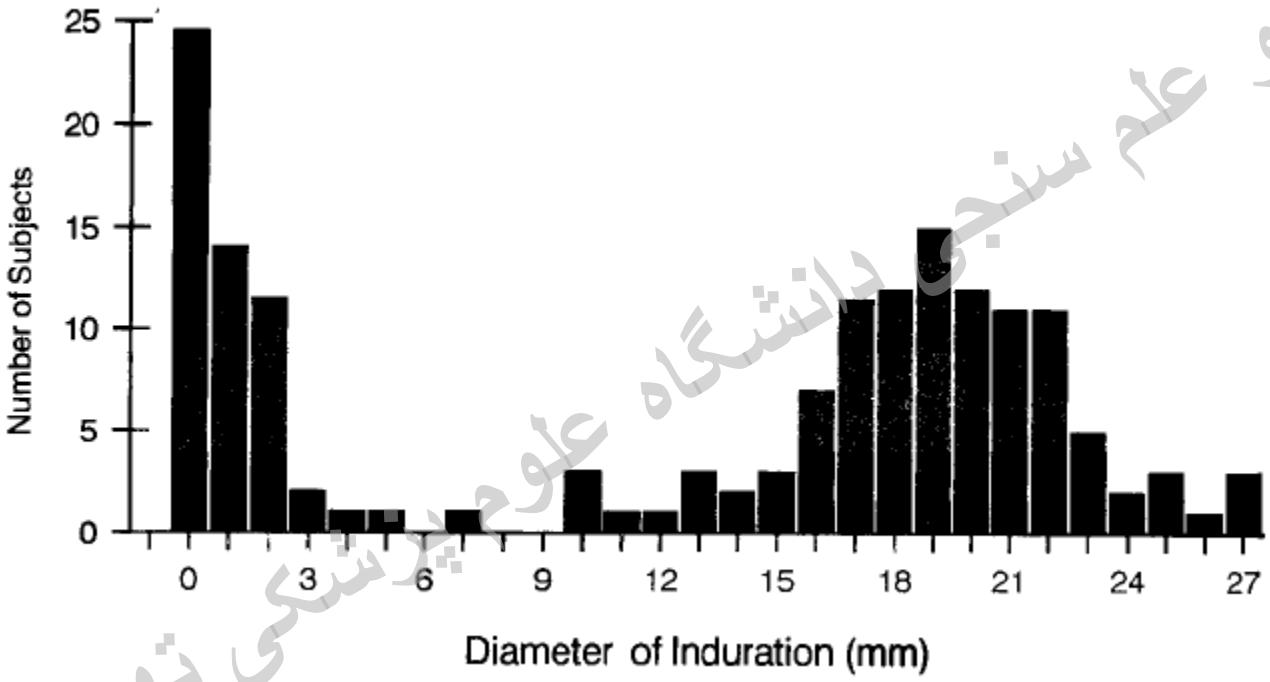


FIGURE 5–1 ▼ Distribution of tuberculin reactions. (Adapted from Edwards LB, Palmer CE, Magnus K: BCG Vaccination: Studies by the WHO Tuberculosis Research Office, Copenhagen. WHO Monograph No. 12. Geneva, WHO, 1953.)

Continuous Measurements

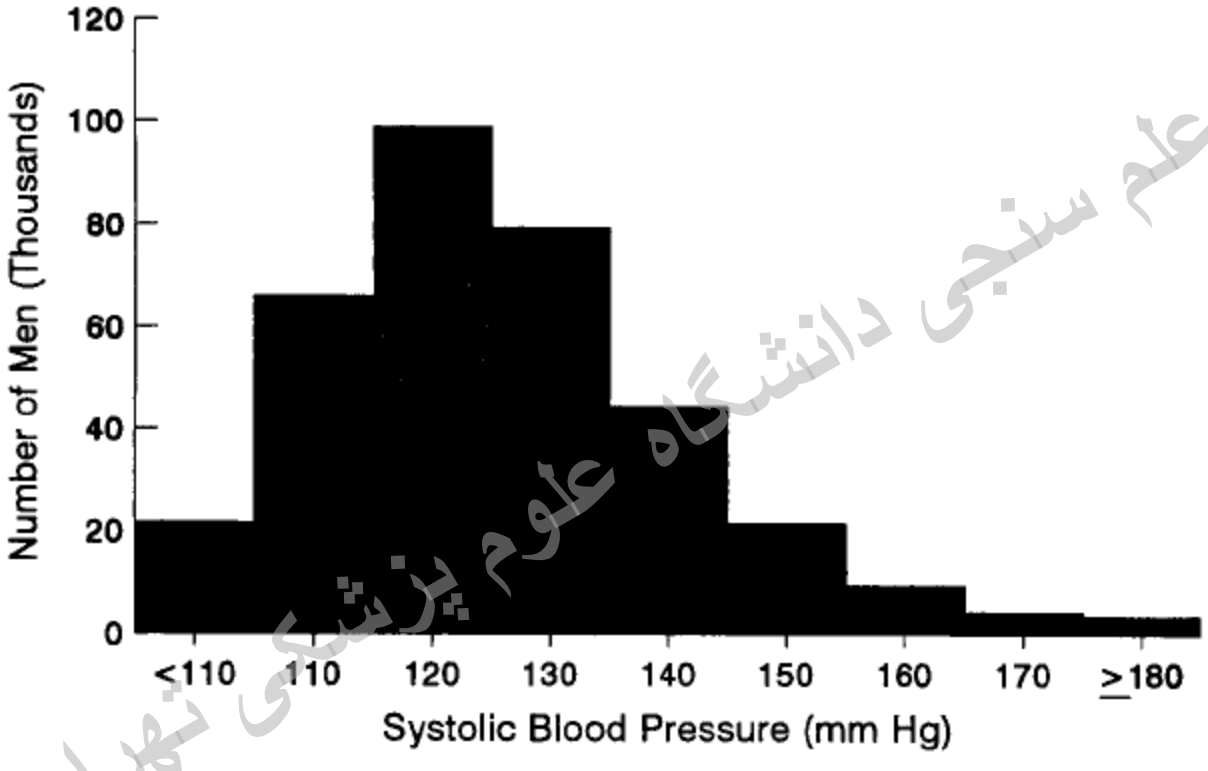
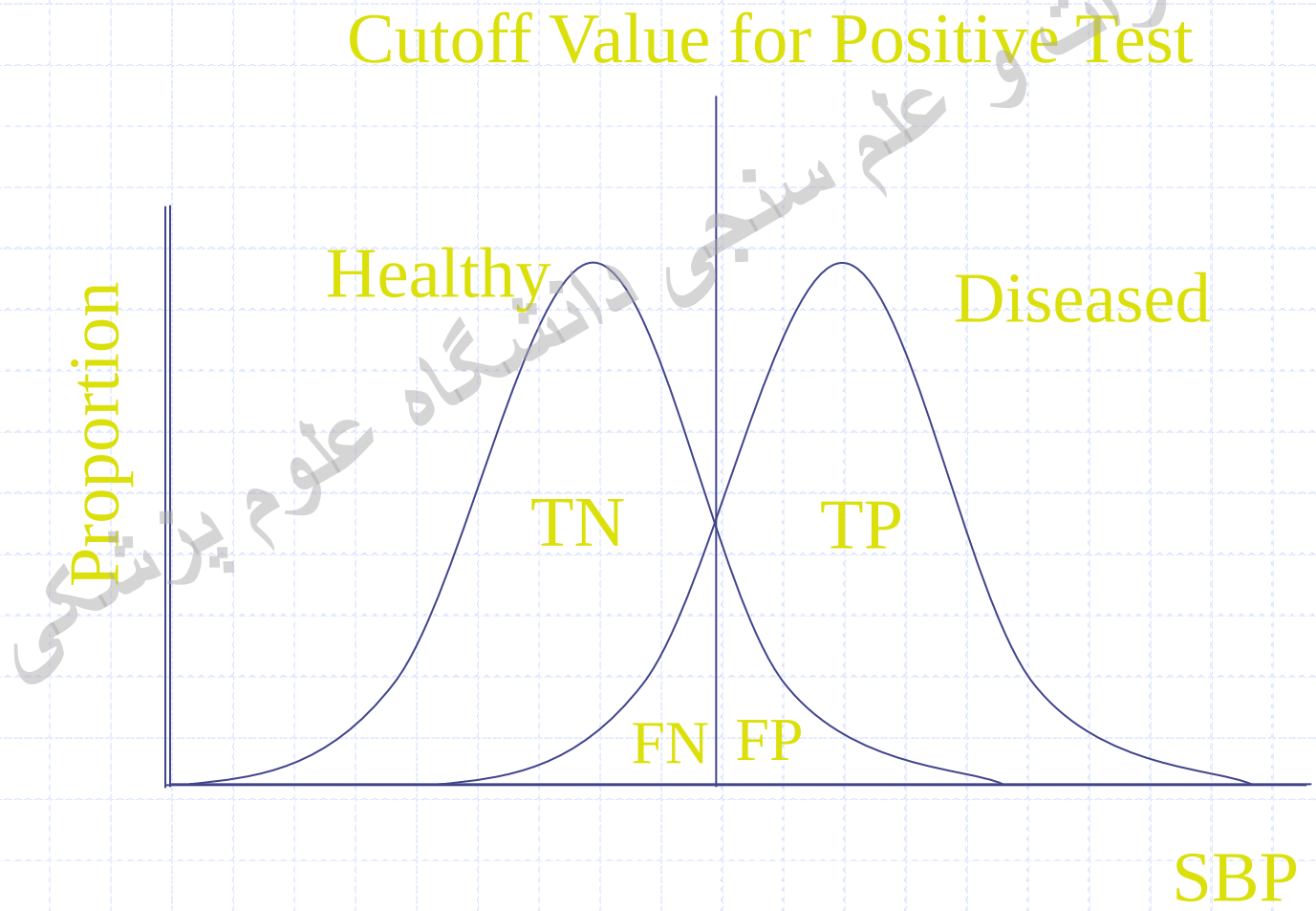
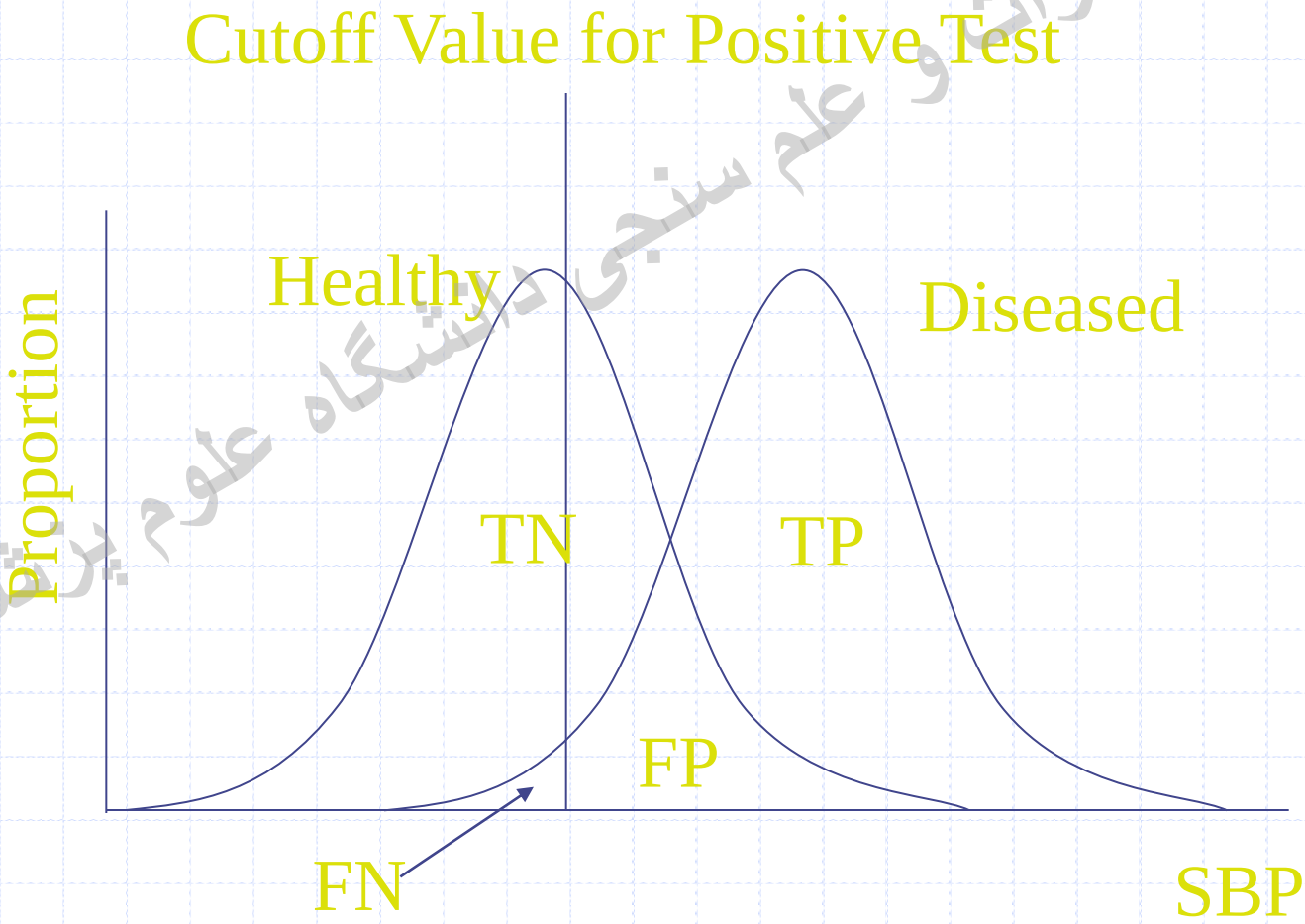


FIGURE 5–2 ▼ Distribution of systolic blood pressure for men screened for the Multiple Risk Factor Intervention Trial. (Data from Stamler J, Stamler R, Neaton JD: Blood pressure, systolic and diastolic, and cardiovascular risks: U.S. population data. Arch Intern Med 153:598–615, 1993.)

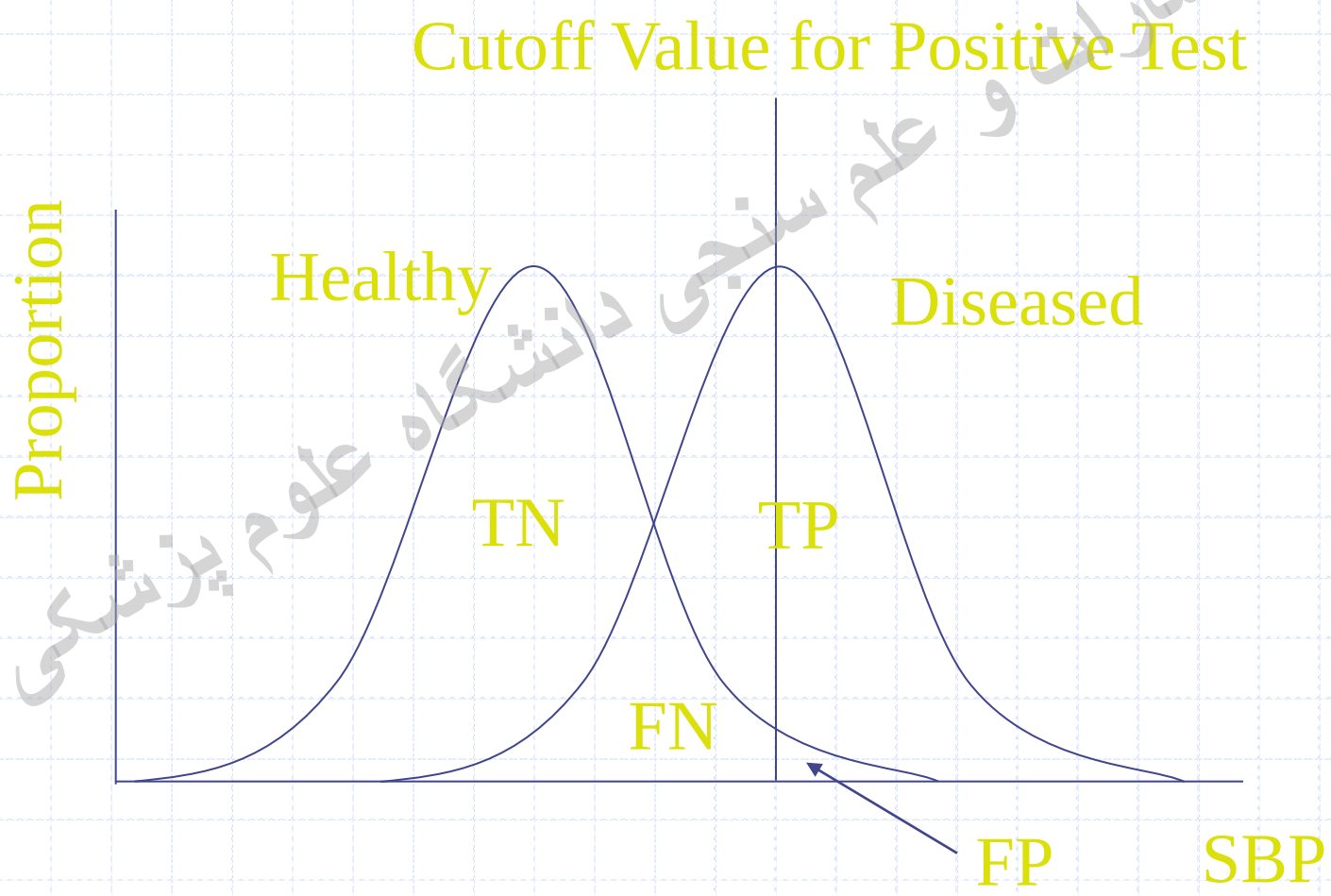
Continuous Measurements



Continuous Measurements



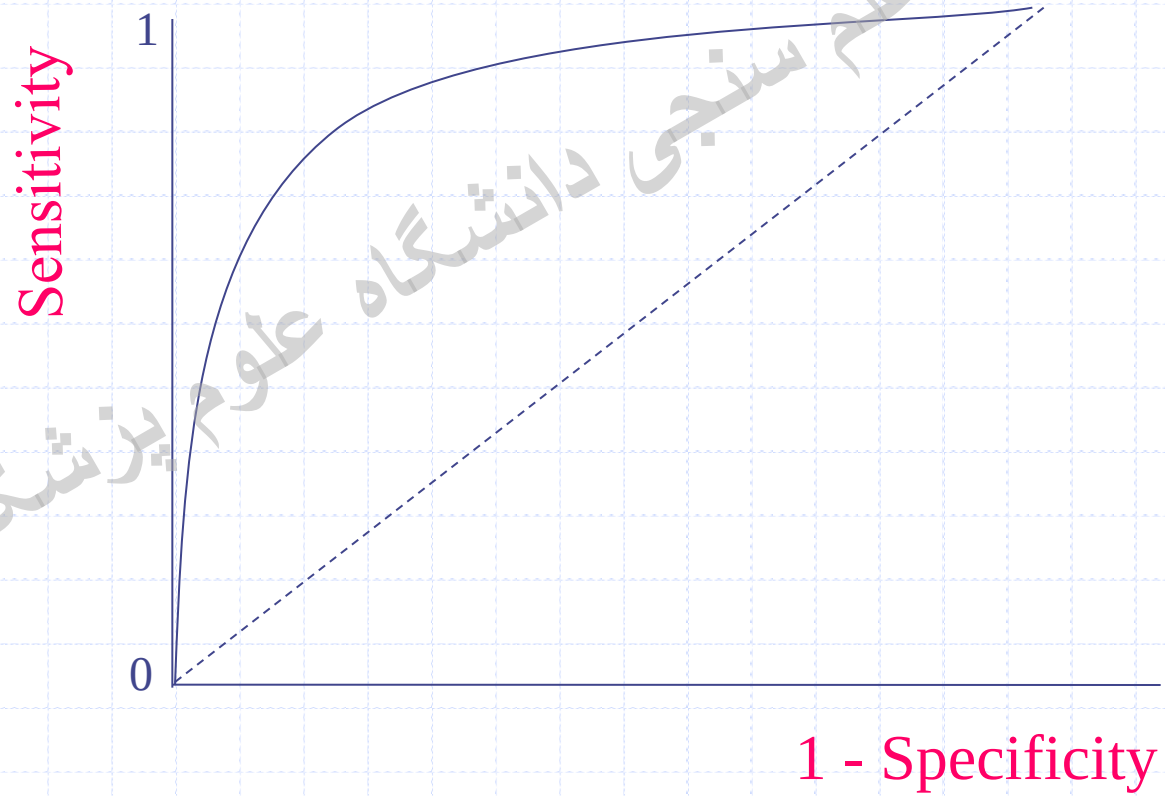
Continuous Measurements



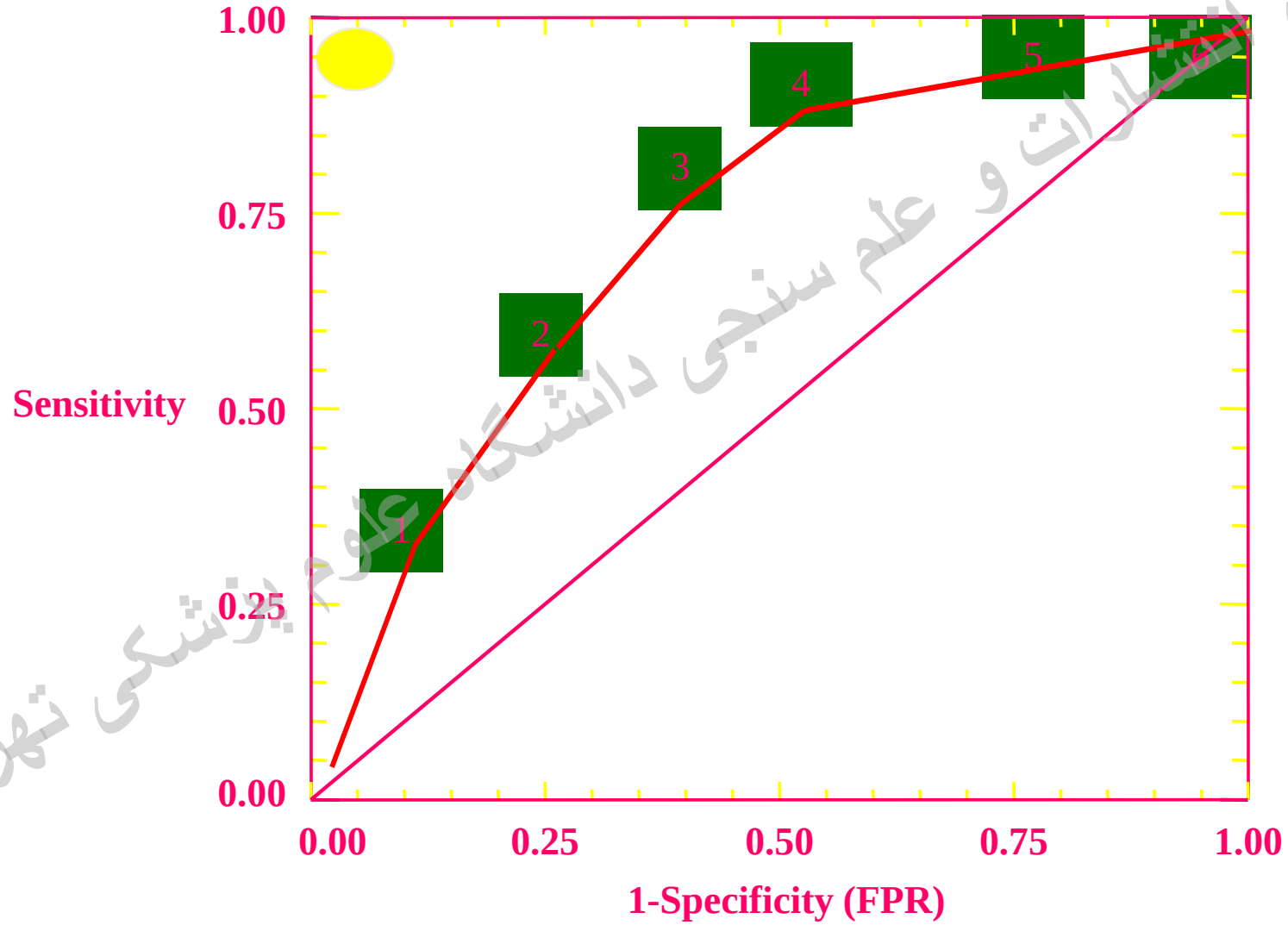
Receiver operator curves

- By plotting the sensitivity and specificity of a test for different cut off points a ROC can be produced which helps illustrate the optimum cut off point to use.

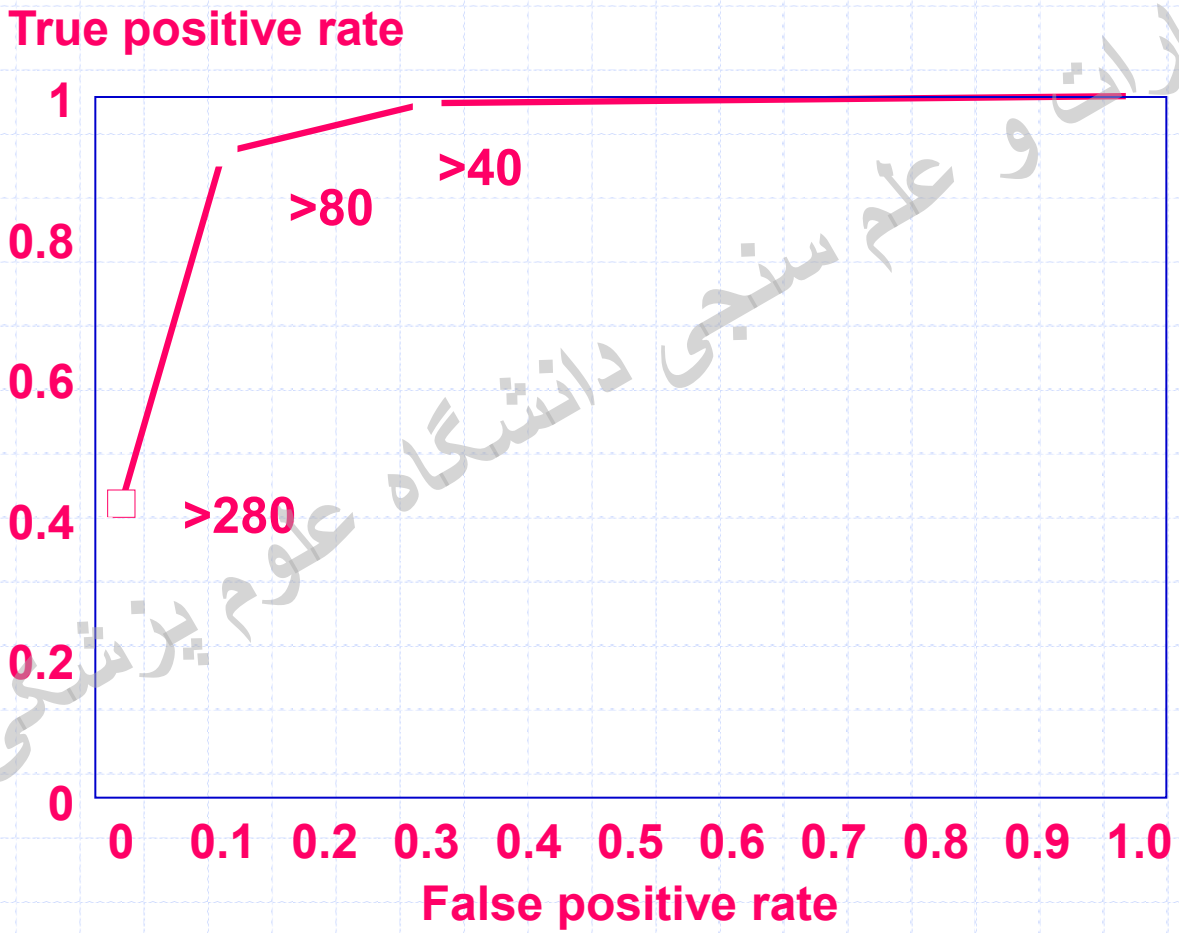
Receiver Operator Characteristic Curve
ROC Curve



ROC Curve Analysis



ROC for creatinine kinase for diagnosing MI



CASP checklist

Biases in diagnostic studies

- Verification bias
- Review bias
- Spectrum bias

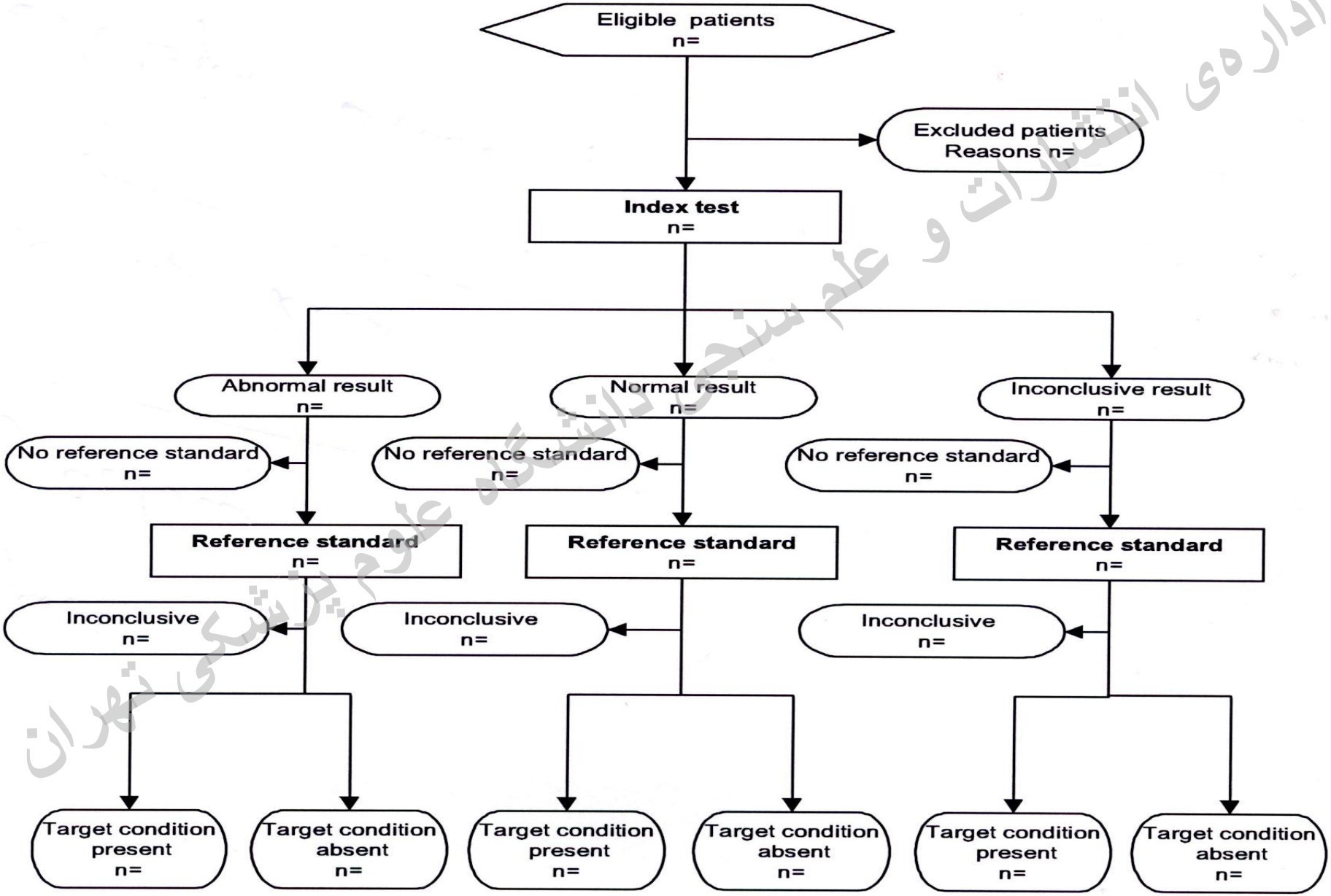
Standards for Reporting of Diagnostic Accuracy (STARD)

Improve the *accuracy* and completeness of research reporting and allow readers to assess the "potential for *bias*" in the study reported.

Always use:

- **FLOW CHART or Diagram**
- **CHECKLIST**

FLOW CHART or Diagram



STARD checklist

Section & Topic	No	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)

Section & Topic	No	Item
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses

Section & Topic	No	Item
METHODS		
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	**6**	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	**10a**	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined

Section & Topic	No	Item
RESULTS		
Participants	**19**	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

CASP

Critical Appraisal Skills Programme (CASP)

making sense of evidence

12 questions to help you make sense of a diagnostic test study

Public Health Resource Unit, England (2006).

Three broad issues

- Are the results of the study valid?
- What are the results?
- Will the results help me and my patients/population?

Screening Questions

- **Was there a clear question for the study to address?**
  - *A question should include information about:*
  - *– the population*
  - *– the test*
  - *– the setting*
  - *– the outcomes*
- **Was there a comparison with an appropriate reference standard?**
  - *Is this reference test(s) the best available indicator in the circumstances?*

Are the results of the study valid?

1, 2. Screening Questions

3. Did **all** patients get the diagnostic test and the reference standard?

4. Could the results of the test of have been **influenced** by the results of the reference standard?

5. Is the **disease** status of the tested population clearly described?

6. Were the **methods** for performing the test described in sufficient detail?

what are the results?

7. What are the results?

8. How sure are we about these results?

Will the results help me and my patients/population?

Consider whether you are primarily interested in the impact on a population or individual level

9. Can the results be applied to your patients the population of interest?
10. Can the test be applied to your patient or population of interest?
11. Were all outcomes important to the individual or population considered?
12. What would be the impact of using this test on your patients/population?

Critical appraisal of

SECONDARY STUDIES

secondary study

- A secondary study does not generate any data from direct measurements, instead, it analyses a set of primary studies and usually seeks to aggregate the results from these in order to provide stronger forms of evidence about a particular phenomenon.

What is a systematic review?

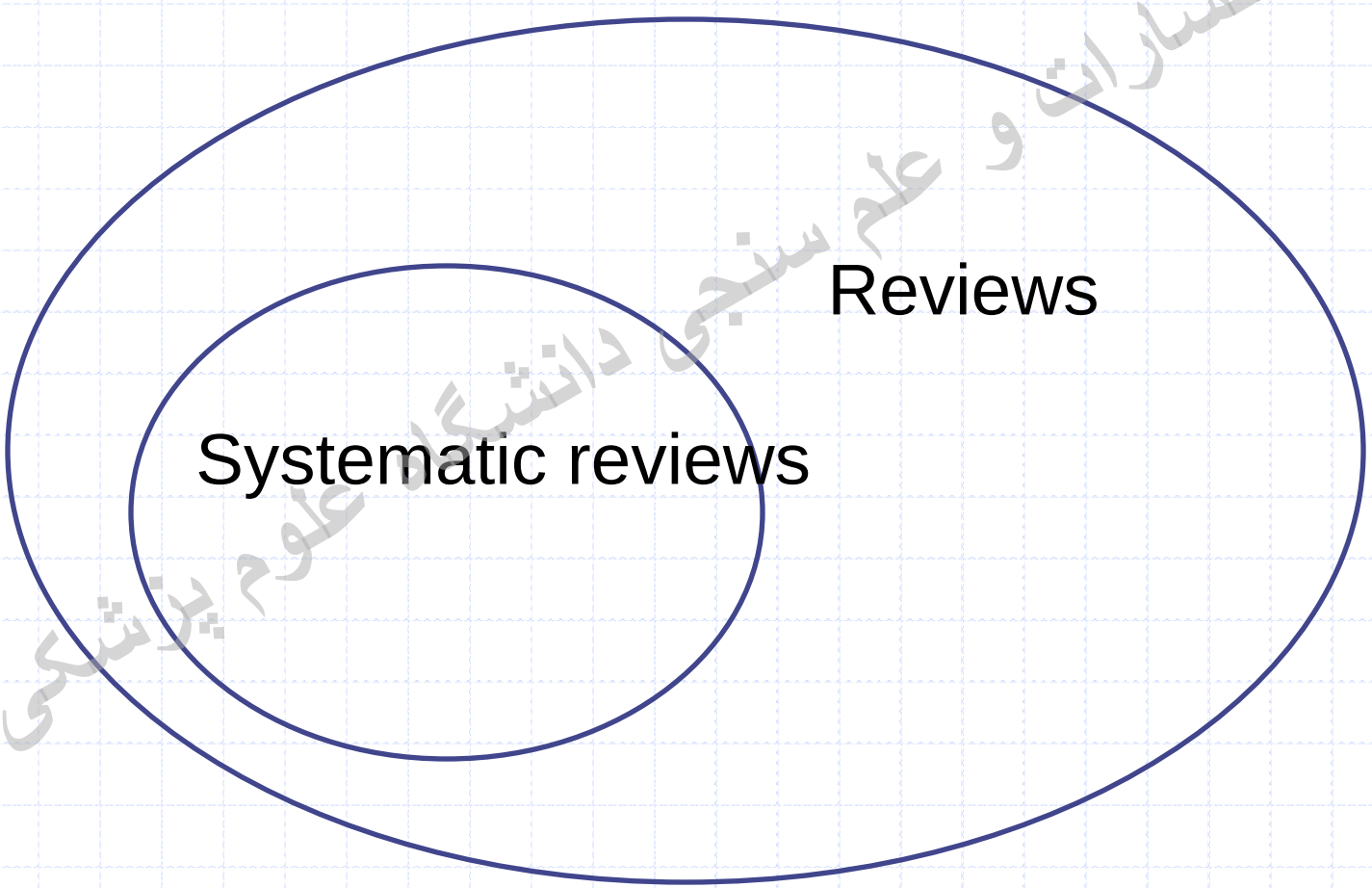
- A review that has been prepared using some kind of systematic approach to minimising biases and random errors, and that the components of the approach will be documented in a materials and methods section

Chalmers et al, 1995

What is a meta-analysis?

- A statistical analysis of the results from independent studies, which generally aims to produce a single estimate of the treatment effect Egger et al, 2001

What is a systematic review



Some of the Appraising tools

Appraising systematic reviews

- Critical Appraisal Skills Program (CASP): Systematic Reviews
- Systematic Review (of therapy) Worksheet
- ARIF (Aggressive Research Intelligence Facility)

Appraising meta-analyses

- QUOROM Statement Checklist

PRISMA Checklist

- The 27 checklist items pertain to the content of a systematic review and meta-analysis, which include the title, abstract, methods, results, discussion and funding.

CASP

Critical Appraisal Skills Programme (CASP)

making sense of evidence

10 questions to help you make sense of reviews

Public Health Resource Unit, England (2006)

Screening Questions

1. **Did the review ask a clearly-focused question?**
2. **Did the review include the right type of study?**
   - *address the review's question*
   - *have an appropriate study design*

Detailed Questions

3. Did the reviewers try to identify all relevant studies?

- *which bibliographic databases were used*
- *if there was personal contact with experts*
- *if the reviewers searched for unpublished studies*
- *if the reviewers searched for non-English-language studies*

Detailed Questions

4. Did the reviewers assess the quality of the included studies?

- *if a clear, pre-determined strategy was used to determine which studies were included.*
  - *a scoring system*
  - *more than one assessor*

Detailed Questions

5. If the results of the studies have been combined, was it reasonable to do so?

– *the results of each study are clearly displayed*

– *the results were similar from study to study*

(look for tests of heterogeneity)

– *the reasons for any variations in results are discussed*

Detailed Questions

6. How are the results presented and what is the main result?

- *how the results are expressed (e.g. odds ratio, relative risk, etc.)*
- *how large this size of result is and how meaningful it is*
- *how you would sum up the bottom-line result of the review in one sentence*

Detailed Questions

7. How precise are these results?

8. Can the results be applied to the local

9. Were all important outcomes considered? (*individual, policy makers and professionals, family/caregivers, wider community)*

10. Should policy or practice change as a result of the evidence contained in this review? (*whether any benefit reported outweighs any harm and/or cost. If this information is not reported can it be filled in from elsewhere?)*

THANK YOU