

# Statistics as a methodology for calculating clinical uncertainty. What kind of studies are relevant to practice?

António Vaz Carneiro, MD, PhD, FACP, FESC

Institute for Evidence Based Health

University of Lisbon School of Medicine

## Introduction

By the beginning of the 21st century, the physician is expected to practice as much as possible based on the best available scientific evidence.

In a progressively more restrictive environment in terms of health resources, it is desired by patients, health managers, insurers, and policy makers that interventions available to treat patients have convincing evidence of efficacy and safety. In addition, these stakeholders also expect all of this to be done at a reasonable and affordable cost. Thus, when applying evidence to practice, the benefits of clinical work should be specified, the cost to be calculable and - more difficult - the evidence itself to be definable and measurable.

The science that should underlie clinical practice comes from clinical studies of good quality. By "clinical" we mean studies that seek to answer patient-relevant questions in patient samples with clinically important patient indicators, and whose results are applicable to patients similar to those in the study.

Clinical questions need studies of good quality that can generate credible answers. Different questions imply different study designs: a therapeutic intervention is best answered by a prospective randomized controlled trial (RCT), but if one wants to define the diagnostic properties of a new test compared to another, then the design is a cross sectional study. This methodological approach is not intuitive, requiring for

each type of study detailed information on its indications, advantages and disadvantages.

## **Causality in medicine**

Determination of causality in medicine is essential, regardless of the context of clinical practice. For example, when a doctor decides on a particular treatment regimen, he assumes that it will cause an improvement in the patient's medical condition. Similarly, when the clinician identifies any risk factor and proposes to modulate or eliminate it, it is because he is convinced that it is related to the onset of the disease, or that it may cause it directly. The key question is therefore whether the observed association between a causal factor (intervention or exposure) and a particular outcome/effect is a cause-effect relationship: if so, can action be taken accordingly?

We can then start by defining what is considered a cause: a factor is considered causal if its operation increases the frequency of a result/effect. It is important to remember here that a preventive factor may also be causal, but operating in reverse, decreasing the frequency of a result/effect.

The medical concept of causality is complex because a disease is usually caused not by one but by several etiological factors working together: tobacco causes lung and bladder cancer, chronic obstructive pulmonary disease, peptic ulcer disease and coronary disease (CD), but CD has several causes besides tobacco (hypercholesterolemia, diabetes, hypertension, genetic burden, etc.).

Determining causality in medicine is very difficult. Indeed, it is impossible: the best that can be achieved, when there is high-quality empirical experimentation data to support it, is to increase our belief in a cause-effect relationship. Thus, once convinced of the quality of the baseline evidence, we come to believe in the causality of the studied factor. Of course we can also take the opposite process, that is, try to obtain evidence that a factor is not the cause of a particular effect - the reasoning is similar.

One of the first problems in establishing a cause-effect relationship is to differentiate between association and causality. It is obvious that the factor and the effect must be

associated if there is a causal relationship between them, but not all associations are causal: there is an association between lighter possession and CD, for example, but CD is not caused by that: the association exists because smokers need lighters, and smoking is a major cause for CD.

The table I shows a possible set of guides that can be helpful to establish causation with a significant degree of certainty.

Table I - A scheme for assessment of causation (1)

<b>A. DESCRIPTION OF THE EVIDENCE</b>
What was the exposure or intervention?
What was the outcome?
What was the study design?
What was the study population?
What was the main result?
<b>B. INTERNAL VALIDITY: CONSIDERATION OF NON-CAUSAL EXPLANATIONS</b>
Are the results likely to be affected by observation bias?
Are the results likely to be affected by confounding?
Are the results likely to be affected by chance variation?
<b>C. INTERNAL VALIDITY: CONSIDERATION OF POSITIVE FEATURES OF CAUSATION</b>
Is there a correct time relationship?
Is the relationship strong?
Is there a dose-response relationship?
Are the results consistent within the study?
Is there any specificity within the study?
<b>D. EXTERNAL VALIDITY: GENERALIZATION OF RESULTS</b>
Can the study results be applied to the eligible population?
Can the study results be applied to the source population?

Can the study results be applied to the relevant populations?
<b>E. COMPARISON OF THE RESULTS WITH OTHER EVIDENCE</b>
Are the results consistent with other evidence, particularly evidence from studies of similar or more powerful study design?
Does the total evidence suggest any specificity?
Are the results plausible in terms of biological/behavioral mechanism?
If a major effect is shown, is it coherent with the distribution of the exposure and the outcome?

Causal relationships are absolutely fundamental in medicine and exist in all areas of health (table II). This means that the systematic scientific approach to produce high quality evidence is central to support decision-making at all levels of the health systems (2).

Table II - The relevance of causal relationships

Field of practice	Goals
Therapy (clinical care)	Improvement in patient's condition
Health services (management)	Improvement in health of community
Health policy (government)	Resource allocation
Medical education (teaching)	Improvement in learning and practice

## Types of clinical studies and their measures of association

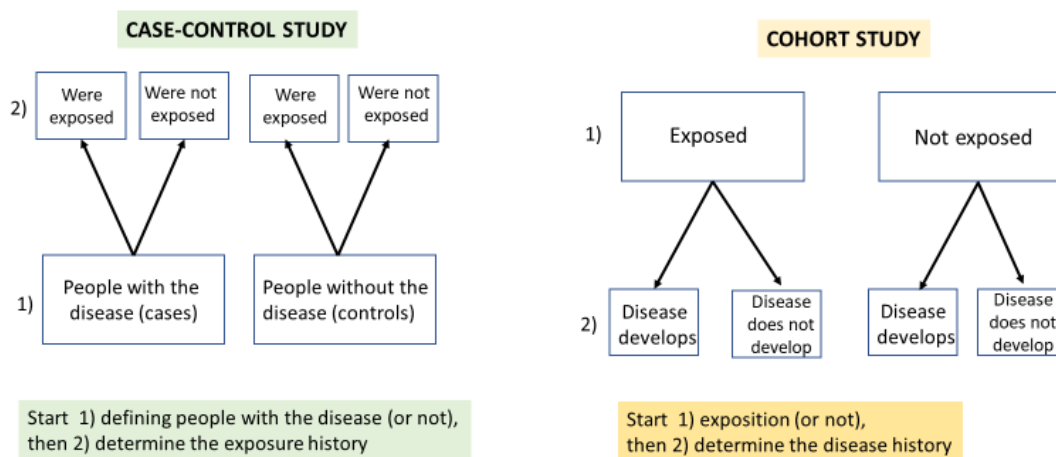
The definition of causal factor (again, a factor is considered causal if its operation increases the frequency of a result/effect) allows one to select the type of studies that prove causality. Indeed, the definition implies that: 1) the individuals presenting the outcome should have a higher frequency of past exposure to the causative agent or 2)

patients affected by the causative agent will have an increased frequency of the outcome.

Therefore, comparative studies can be of two types: those testing implication #1, comparing one group of patients who have already experienced the outcome with another group where the patients have not experienced an outcome, retrospectively identifying the causal factor (case-control study) and those testing implication # 2, comparing one group of patients exposed to the putative causal factor with another group of unexposed patients and follow them during a specific period of time to look for outcomes (cohort study). Of course, what is sought here is the establishment of the quantitative causal relationship, where the factor is neither necessary nor sufficient for the onset of the effect, with the results expressed in risks, ratios, etc.

In the figure 1 there is a graphic format of the differences between this two types of typical causation studies (3).

Figure 1 - Comparison between case-control and cohort studies



Cohort studies may be descriptive (when limited to describing the incidence of certain outcomes over a period of time) and analytical (when analysing associations between predictive factors and outcomes); they can be prospective, when they begin with factor exposure and establish future follow-up over a period of time or retrospective,

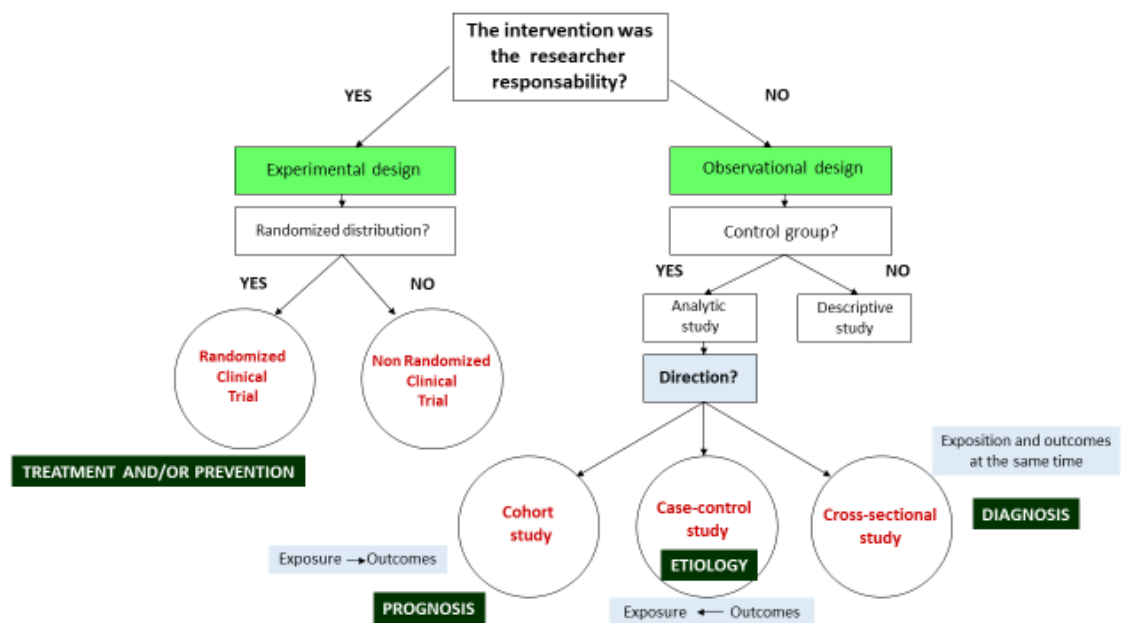
when they use past information to identify factor exposure and provide follow-up information since then.

In intervention cohort studies (experimental) - of which the prospective randomized controlled trial (RCT) is the paradigm - researchers randomly control subjects' exposure to putative causal factor (controlled study), defining the possible causal relationship between the two. In observational cohort studies, researchers do not influence the variables present, but simply analyse the relationships present (4).

Each type of study has its advantages and disadvantages so the researchers will decide which study design to adopt to obtain an answer to each individual research question. These questions can be answered with studies that measure and obtain differences in outcomes, called measures of association.

In figure 2 we show the different types of studies (experimental vs. observational) that should be used according to the specific clinical question.

Figure 2 – Types of clinical studies



In table III the different type of the most common studies is presented with its specific measures of association (5).

Table III – Types of studies and their measures of association

TYPE OF STUDY	MEASURE OF ASSOCIATION
Case-control	Odds Ratio (OR)
Cohort	Relative Risk (RR)
Cross-sectional	Prevalence
Clinical trials	Absolute Risk Reduction (ARR), Relative Risk Reduction (RRR) e Number Needed to Treat (NNT)

The calculation of the measures of association can be achieved by inserting into the cells of the table IV (a 2 X 2 contingency table) the number of patients duly classified according to exposure, and presence or absence of the outcome of interest.

This table constitutes the basis of all interpretation of the data obtained in the studies, allowing for translation of data to clinical practice.

Table IV – 2 X 2 table for measures of association

	Outcome +	Outcome -	Total
Exposed	a	b	a + b
Non-exposed	c	d	c + d
Total	a + c	b + d	a + b + c + d

The letters and respective position in the table are there by convention and should be applied, irrespective of the type of study one may want to design.

## Example of diagnostic, treatment and prognostic questions and their studies

In practice there are several possible clinical questions, depending on the area of knowledge one wants to clarify. The most frequent questions are diagnostic, therapeutic and prognostic.

We will give now three different example to support the interpretation of research data to respond to the aforementioned questions. We start with a diagnostic question, followed by a therapeutic one and, finally, a prognostic one.

### **IF THE TEST IS ABNORMAL, HOW LIKELY IS THE PATIENT TO HAVE THE DISEASE?**

The method to define the discriminatory characteristics of a diagnostic test is a cross-sectional study. This implies that there is a comparison between the new test and the gold standard in terms of disease definition (the gold standard being responsible for the background diagnosis).

Referring to the table IV, the results in cell "a" are called true positives (both tests are abnormal), cell "d" true negatives (both tests are normal), cell "b" false positives (the new test is abnormal but, according to the results of the gold standard, the patient does not have the disease) and, finally, cell "c" false negatives (new test is normal but, according to the results of the gold standard, the patient has the disease).

The calculations of measures of association in a diagnostic study are:

- SENSIBILITY =  $a/a+c$
- SPECIFICITY =  $d/b+d$
- POSITIVE PREDICTIVE VALUE =  $a/a+b$
- NEGATIVE PREDICTIVE VALUE =  $d/c+d$
- ACCURACY =  $(a+d) / (a+b+c+d)$
- POSITIVE LIKELIHOOD RATIO =  $a/(a+c) / b/(b+d)$
- NEGATIVE LIKELIHOOD RATIO =  $c/(a+c) / d/(b+d)$

#### PRACTICAL EXAMPLE (Ann Intern Med 2003;138:787-794)

*It is important to diagnose deep vein thrombosis (DVT), because this condition has an important local morbidity and a potential high mortality from pulmonary embolus. One of the tests available to diagnose DVT is compression ultrasonography (USC). This technique is good, but has some problems in terms of reproducibility among operators. So a new blood test - the D-dimer test (easy to perform and without operator*



problems), was compared with the compression ultrasonography in 556 patients suspected of having DVT.

The results were the following:

- 55 patients with DVT had an abnormal D-dimer test (DD +)
- 1 patient with DVT had a normal D-dimer test (DD -)
- 198 patients without DVT had DD + and
- 302 patients without DVT had a DD -

Replacing these values in the table, it looks like this:

	USC + (disease present)	USC – (disease absent)	Total
DD +	55	198	253
DD -	1	302	303
Total	56	500	556

Now, one can calculate the measures of association of this study:

- Sensibility =  $a/a+c = 55/56 = 0,98 = 98\%$
- Specificity =  $d/b+d = 302/ 500 = 0,60 = 60\%$
- Positive predictive value =  $a/a+b = 55/253 = 0,22 = 22\%$
- Negative predictive value =  $d/c+d = 302/303 = 0,9967 = 100\%$
- Accuracy =  $(a+d) / (a+b+c+d) = 357/556 = 0,64 = 64\%$
- Positive likelihood ratio =  $a/(a+c) / b/(b+d) = 0,98/0,4 = 2,45$
- Negative likelihood ratio =  $c/(a+c) / d/(b+d) = 0,017/0,6 = 0,03$

How should we interpret clinically these numbers (please do not forget that this study should support clinical care)?

- Only 22 out of 100 d-dimer + patients had DVT
- Almost no patient with negative d-dimer had DVT
- 64 out of 100 patients were correctly classified using the d-dimer test

- A patient with an abnormal d-dimer result has 2.5 more chances of having DVT than one with normal d-dimer.

### IF I PRESCRIBE THIS MEDICINE, HOW LIKELY IS IT THAT THE PATIENT GETS BETTER?

In a study of a therapeutic intervention (drug therapy, surgery, etc.) or a preventive measure (vaccination, smoking cessation, etc.), the design of the study to prove efficacy and calculate harm is a randomized controlled trial (RCT).

The typical pharmacological RCT is called a parallel study because it randomizes patients to 2 groups: the one that receives the test drug (called the experimental group) and the one that receives a placebo or a comparator drug (control group). These patients are then followed prospectively during a pre-determined period, and measures of association – as already indicated ARR, RRR and NNT - are then determined.

The calculations of measures of association in an intervention study (RCT) are, using Table IV (6):

- ABSOLUTE RISK (AR) =  $[a / (a + b)]$  and  $[c / (c + d)]$
- ABSOLUTE RISK REDUCTION (ARR) =  $[c / (c + d)] - [a / (a + b)]$
- NUMBER NEEDED TO TREAT (NNT) =  $(1 / RRA) \times 100$
- RELATIVE RISK (RR) =  $[a / (a + b)] / [c / (c + d)]$
- RELATIVE RISK REDUCTION (RRR) =  $[c / (c + d)] - [a / (a + b)] / [c / (c + d)]$

#### PRACTICAL EXAMPLE (Lancet 1999; 353: 9–13):

*In the CIBIS II trial, bisoprolol (a beta-blocker drug) was administered as an add-on during 1.3 years to 1327 patients with congestive heart failure (CHF). These patients were compared with 1320 similar subjects doing placebo. The main endpoint was all cause mortality. The trial was stopped early because all-cause mortality was significantly lower in the bisoprolol group (156/1327 patients, 12%) than in the placebo group (228/1320 patients, 17%) ( $p < 0.0001$ ).*

*Making the direct calculations, the ARR was 5% (17-12%), the RRR was 29% (17-12/17%) and the NNT was 20 (1/5 x 100).*

This is a very effective therapeutic scheme, because you only have to treat 20 patients with CHF to save a life.

### **IF THE PATIENT HAS THIS RISK FACTOR, HOW LIKELY IS HE TO HAVE THE DISEASE IN THE FUTURE?**

As already said, a risk factor is a cause of an event if its operation increases the frequency of the event.

One of the best methodologies to investigate risks is the prospective cohort study: in it, a group of people is selected, none of whom has experienced the outcome of interest, but all of whom could experience it. For each risk factor, subjects of the cohort are classified either as exposed (i.e., possessing the factor in question) or unexposed. They are all followed up over time to see which of them experience the outcome, and the rates of the outcome events are compared in the exposed and unexposed groups.

#### PRACTICAL EXAMPLE (1):

*In a prospective study looking to the relationship of smoking with lung cancer, a group of 21,624 subjects were followed for 10 years. At the end of the study, of the 10,620 smokers 120 had a diagnosis of lung cancer. Of the 11,004 non-smokers, only 4 presented with lung cancer.*

The calculations of measures of association in a prognostic study are as follows (7):

- Among the exposed individuals, the risk of achieving the outcome is defined as:  $R(\text{exposed}) = \text{Exposed with outcome} / \text{all exposed} = a / (a + b)$ . The risk may then vary from 0 (none of the exposed has the outcome) and 1 (all exposed have the outcome)
- Concerning the unexposed individuals, the risk of achieving the outcome can be defined as:  $R(\text{unexposed}) = \text{not exposed to outcome} / \text{all non-exposed} = c / (c + d)$

- Relative risk:  $RR = R(\text{exposed}) / R(\text{unexposed}) = a/(a + b) / c/(c + d)$ . The value of RR tells us how many times more individuals with the risk factor come to develop the outcome compared to non-smokers.

	Outcome +	Outcome -	Total
Exposed	120	10500	10620
Non-exposed	4	11000	11004
Total	124	21500	21624

- $R(\text{exposed}) = a / (a + b) = 120/10620 = 0.0113$
- $E(\text{unexposed}) = c / (c + d) = 4/11004 = 0.00036$
- Relative risk:  $RR = R(\text{exposed}) / R(\text{unexposed}) = 0.0113 / 0.00036 = \mathbf{31}$

The value of RR tells us that is 31 times more likely that individuals who smoke regularly will develop lung cancer, compared to non-smokers.

As one can see, this result gives us only the dimension of the harm done by smoking, not the basal (absolute) harm rates. But it is useful to calculate the impact: after all, a RR of 1.2 or 1.3 is considered to be clinically significant. Here we have a RR of more than 3000% (that can be considered almost directly causal).

## Conclusions

Determination of causality in medicine is of great importance, both in the field of diagnosis as well as in therapy and prevention (to mention only these).

The relationship between the putative causal factor and the observed outcome/effect - which can be classified as necessary, sufficient, both or neither - must be known. In medicine, most situations fall into the latter type: the action of the causal factor increases the frequency of the effect, but this does not always occur, or occurs in the absence of the causal factor.

The determination of causality excludes the existence of only one association, and the clinician may use a set of practical rules for the effect (temporality, strength, dose response, reversibility, consistency, biological plausibility, specificity and analogy).

Better is to conduct studies specifically designed to prove causation, which include case-control, prospective and retrospective cohort studies, cross-sectional and clinical trials (among others). The results, when well integrated, will increase the quality of clinical care (8).

## **Bibliography**

*Note: due to the specific nature of this field of clinical, statistical and epidemiological knowledge, the bibliography will include only classic books on the subject (as opposed to individual papers).*

- 1) Mark Elwood. **Critical Appraisal of Epidemiological Studies and Clinical Trials**. 4<sup>th</sup> Edition, 2017, Oxford University Press
- 2) Kameshwar Prasad. **Fundamentals of Evidence-Based Medicine**. 2<sup>nd</sup> Edition, 2013, Springer
- 3) **Leon Gordis**. **Epidemiology**. 4<sup>th</sup> Edition, 2017 Saunders
- 4) Amit Kaura. **Crash Course Evidence-Based Medicine: Reading and Writing Medical Papers**. 2013, Mosby Elsevier
- 5) Richard K. Riegelman. **Studying a Study and Testing a Test. How to Read the Medical Evidence**. 6<sup>th</sup> Edition, 2013, Lippincott Williams and Wilkins
- 6) **Medical decision making**. Sox HC, Higgins MC, Owens DK Editors. 2<sup>nd</sup> Edition, 2013, Wiley-Blackwell
- 7) Trisha Greenhalgh. **How to Read a Paper. The Basics of Evidence-Based Medicine**. Th Edition 2014, Wiley - BMJ Books
- 8) **User's Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice**. Guyatt G, Rennie D, Meade MO, Cook DJ Editors. 3<sup>th</sup> Edition, 2015, JAMA Evidence, McGraw Hill