CHAPTER XII

THE KEY POINTS OF THE INTERNAL VALIDITY OF A STUDY BIAS AND CONFONDING FACTORS.

P Duhaut, J Schmidt

All research methods in clinical epidemiology (or in evidence-based medicine) consist in trying to establish the principles governing our clinical practice on observed, quantifiable, measured, verified facts, rather than on theoretical assertions, of the simple empiricism, even widespread medical beliefs that have, in one way or another, taken root in our knowledge or our attitudes. This effort has undoubtedly led to significant progress. No method of investigation or observation is perfect, however, and the literature is full of contradictory or different results. Clinical epidemiology, in trying to identify patterns, does not establish 'hard' truths, such as the discovery of DNA and its role in genome transmission. This chapter reviews some of the interpretative limitations of the studies.

Clinical Epidemiology, whose methods are partly inspired by classical Epidemiology methods, has been developed to help the clinician answer the many questions related to his daily practice in the fields of prevention, diagnosis, treatment, evaluation of prognosis, nosology...:

What are the risk factors for colon cancer? What is the best way to prevent complications of high blood pressure? What is the incidence or prevalence among the elderly population? What is the best treatment for myocardial infarction taken in the first hours? What is the best therapeutic strategy for stage II breast cancer? What is the optimal diagnostic approach in the face of a long-term fever? What is the sensitivity and specificity of second-generation serological tests for hepatitis C?...

Each doctor, according to his experience, his therapeutic habits, his school of training, can have an opinion based on bases that seem reasonable to him... Opinion often different from that of other experienced doctors, even diametrically opposed.

Clinical studies seek to resolve these questions or these contradictions by quantifying the facts on a reproducible basis from a sufficiently large number of patients to obtain a statistically significant response. The result must - in principle - be applicable to the next patient suffering from the same pathology.

Methodological tools for epidemiological and statistical analysis have been developed for this purpose. We are forced to note, however, that there are differences in conclusion or contradictions between many apparently well-conducted studies, as glaring as those between individual opinions.

The origins of these differences are many:

- First of all, they are due to a possible lack of precision in a correctly constructed study.

- They can then correspond to the presence of hidden defects, called biases, more or less inherent in the structure of the studies in question. The biases condition the internal validity of the study.
- Thirdly, they depend on the characteristics of the populations studied: the same disease can have different causes, and the same treatment can have variable efficacy, in different populations. This introduces the notion of external validity, or ability to be generalized, of the study.
- Finally, they call on knowledge of the limits of applicability of the epidemiological and statistical tool to the analysis and understanding of a phenomenon.

These questions resolved - if it is possible - leave the problem of the application to an individual of results obtained at the level of a group. A particularly important problem for the clinician, who often has to refer to group studies to determine the cause of an individual's illness, choose the treatment or infer the prognosis... and the paradox of group studies, whose purpose first and ultimate is often to help the clinician to answer the questions he is asking about the individual.

I- CONCEPT OF PRECISION; RANDOM ERROR

Let's imagine a randomized trial seeking to compare a protocol A and a protocol B of chemotherapy in epidermoid cancers of the pharynx (excluding poorly differentiated cancers linked to the Epstein-Bar Virus of the nasopharynx), conducted in two different oncology units I and II .

Suppose these two oncology units are open to any patient likely to present with ENT cancer, without distinction of gender, origin, age, social background, exposure to a particular risk factor or any other factor. All new patients with squamous cell carcinoma hospitalized during the study period were included and randomized between arms A and B on a double-blind basis, eliminating in principle any selection bias. Unit I included 23 patients and Unit II 19. The data, presented below, are analyzed in each department separately.

	Unit 1		Unit 2	
Protocol	A	B	A	B
Patients included	12	11	8	11
Survival at 1 year	6	4	3	5
Survival at 1 year (%)	50%	36%	37%	45%

These two trials, identical, correctly designed, without obvious bias, involving the same type of patients diagnosed in the same way, seem to give contradictory results. What could be the reasons?

A first explanation can be given by random sampling: although there is no selection bias at the level of inclusion in the study (since all the patients were included and the two units are open to all patients) or in one arm (since the allocation was randomized), patients from Unit I or Unit II represent only two randomly drawn samples from the entire patient population with ENT squamous cell cancer worldwide. These two samples may by chance not be equivalent to each other, or may not be equally representative of the general population. They may therefore respond differently to treatment A or B without this being linked to the treatment itself: the real disparity observed between Unit I and Unit II in the effectiveness of the

treatment will in fact correspond to the difference not known, occurred at random sampling, between the two groups.

We can now assume that the difference observed is not real, is not significant. Only a small number of patients were included, and the differences of 14% in one direction and 12% in the other observed in Units I and II are only due to one individual each time: an additional survival occurred by chance in arm B of unit I would increase the success of B to 45%, and an additional survival occurrence in arm A of unit II would increase the success of A to 50%, thus homogenizing the results of the two units ... and eliminating any difference between treatment A and B, in a concordant manner in both centers!

Results obtained from a small number of patients arouse some skepticism. There are, however, statistical "safeguards", expressed in the form of confidence intervals, making it possible to "quantify" the chance in obtaining the results and consequently their significance. It is generally accepted that results are significant when they have less than a 5% chance of being due to chance. Thus, we call the error that is always possible to conclude that a significant difference does not really exist: **error alpha**.

In other words, it is most often requested that the ***alpha error be less than 5%.***

Now let's forget about unit II and assume that 5 patients are alive at one year in arm B of unit I, thus reducing the success of this arm to 45%. There is no longer any notable difference between treatments A and B. Does this mean that the two treatments are of comparable effectiveness (or ineffectiveness)? Or that the study failed to show a difference that actually exists?

A study must include enough patients, be sufficiently powerful, to have an acceptable chance of showing a difference that actually exists. Here again, there is a statistical "safeguard" making it possible to quantify the power of a study, i.e. its chance of arriving at a conclusion reflecting the reality that we are trying to estimate: a good study must have at least an ***80% chance*** of succeeding in resolving the question it asked itself, i.e. ***a power of 80%.*** The error, always possible, of erroneously concluding that there is no difference between the groups being compared because the study failed to show a difference that actually exists, is called **beta error**.

In other words, it is generally required that the ***beta error be less than 20%,*** or that the ***power be greater than 80%.***

Power = 1 - beta.

Sampling chance and results obtained at random from alpha or beta error therefore represent the major sources of error *linked to the lack of precision* of a study that is otherwise correctly constructed and conducted. It is possible, to reduce this risk of error linked to chance, *to increase the size of the sample studied*: this will increase the power of the study, increase the significance of the results, and improve the chances that the sample is truly representative of the population of patients suffering from the disease that it is proposed to study.

In other words, increasing the size of the sample studied decreases the alpha error, decreases the beta error, and decreases the errors related to sampling chance.

A systematic review of trials carried out in anesthesia showed that in 2000, only 56% of randomized trials had sufficient power. The score improved to 86% in 2006, but only 18% of trials with negative results analyzed their risk of beta error [1]. This study reflects the general trend quite well, and it is rare for the power of a study to be considered before concluding, perhaps wrongly, that there is no effect.

II- BIAS

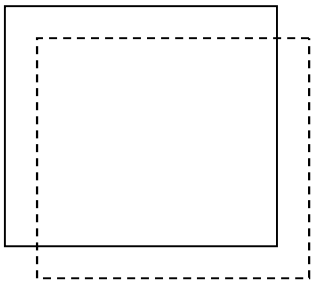
Contrary to the previous errors, the biases are at the origin of errors related to real flaws in the form of the study, appearing at any of the stages of its design. A bias is - in principle - a shortcoming that we must try to reduce to a minimum or eliminate - this is not always possible - before starting the study in the field. A bias is not due to chance.

The defects of form are numerous and certain types of studies predispose more particularly to such or such type of bias. Olli Miettinen proposed a classification into three main categories:

1- Selection bias
2- Biases in measurement, or information, or misclassification
3- The confounding elements

1- Selection bias:

Selection bias arises when the population actually studied is not representative of the population that we wanted to study, and to which we would like to be able to apply the results.



_____: *population we want to study (target population)*

----------: *population finally reached, slightly or heavily different from the target population*

Several factors may contribute to producing a selection bias, which may be significant enough to affect the validity of the results:

Inclusion and exclusion criteria may not have been adequate.

- Imprecise, leaving too much uncertainty as to their interpretation and thus allowing, within the framework of a multicenter study, the inclusion of a heterogeneous population and different samples according to the participating centers.
- Too restrictive or too precise, on the contrary, excluding a large number of subjects from the study and thus jeopardizing the external validity of the study: the results cannot be applied to the unfamiliar patient suffering from the pathology studied, because too different from the "ideal model" on which the study was really based.

In a case-control study, the sensitivity and specificity of the tests or diagnostic criteria for the disease studied are decisive for the selection of cases and controls: too sensitive, they will lead to the inclusion, in the group of cases, of patients without the disease studied. Too specific, they will result in the exclusion of patients who actually have the disease. However, good specificity is preferable to good sensitivity for the validity of the study [2].

The selection of controls in a case-control study represents a very important element of the construction of the study, the results of which are based on the comparison of the group of patients and the group of controls. The ideal choice of controls is made by drawing lots from the population from which the cases come (see Chapter V). This is not always possible. Thus, many studies recruit controls among the family, friends or neighborhood of the cases.

The non-response bias can represent a significant selection bias, even though the previous stages (choice of inclusion and exclusion criteria, choice of the control population in a case-control study) avoided the pitfalls. It can occur when a certain number of patients approached to enter the study do not answer the mail or the telephone call inviting them, or refuse to take part in the study after having been informed.

Non-responding subjects or subjects refusing to participate in the study may be different from the subjects finally included, who will therefore no longer be representative of the entire initially targeted population. Some authors have focused on studying non-responders and have effectively highlighted socio-demographic differences that are significant enough to modify the results of the study undertaken [3].

Possible selection biases can be obvious or more hidden:

Examples:

- The response rate to the Richard Doll questionnaire on the prevalence of excessive alcohol consumption and its potential complications in the cohort of British doctors was 73%, versus 95% for the questionnaires focusing on the complications of smoking. The reduction of 22% undoubtedly corresponds to a significant selection bias, the non-responders for this socially sensitive question being undoubtedly different from the responders. Consequently, the measurement of the complications of the risk factor is biased and the results of the study difficult to interpret, at least on the quantitative level [4].

- The selective survival bias constitutes another possible selection bias: a treatment can of course only be applied to living patients. This selective surviving sample, insofar as the time 0 of the onset of a disease is rarely known, may constitute only the surviving fraction of the population initially affected, and therefore the fraction with the best prognosis. A recent study sought to quantify this bias in the expression of the results of the care of patients admitted to hospital for myocardial infarction: it was able to show that the survival bias increased with the time to admission to hospital, and that the measured 'efficacy' of an objectively ineffective treatment could, in terms of mortality reduction, increase from 4 to 27% if the changing nature of the disease as a function of time spent was ignored [5]: the survivors, by definition, have a better prognosis than patients who died quickly, and this better prognosis may wrongly be attributed to the treatment. This bias can easily arise in any case-control or cohort study: for example, mortality in the population-based Cardiovascular Health Study (CHS) was 40% lower than that observed in several cohorts of registered patients in Medicare-type health insurance programs in the United States, whereas CHS patients were on average older

and the proportion of men was higher [6]. Randomizing a particular population for a therapeutic trial does not make it go away, and various selection biases can still influence trial results [7].

When the inclusion of controls by drawing lots is not possible, it is generally better to prefer controls recruited in the vicinity of the patient to direct friends: they are more likely to come from the population from which the cases come, without however sharing all the habits or risk factors of the cases as friends might do (particularly in terms of smoking or alcohol consumption) [8].

Detection bias is another example of selection bias, which is often very difficult to highlight or quantify.

3 examples:

- A bias of possible detection has been discussed at length in the various case-control studies focusing on the relationship between estrogen replacement therapy in menopause and the occurrence of endometrial cancer (see chapter V) [9], with contradictory results between studies showing at the very least that a study designed to avoid bias can be caught in its own trap [10].
- The usefulness of screening for lung cancer has long been debated, and the main studies carried out in Europe or the United States based on the performance of a simple chest X-ray per year have not shown a clear benefit in terms of survival of screened patients. A recent study carried out in Japan, with systematic chest scans, seems to show an improvement in the survival of screened patients. Two unresolved questions remain open: that of an early cancer detection bias simply lengthening the observation time of patients, and therefore the observed survival, and that of a simple detection bias of small less aggressive cancers not diagnosed on simple chest X-ray, and whose prognosis, potentially better, modifies the survival data previously collected on more serious patients [11].
- The administration of finasteride, a treatment for benign prostatic hyperplasia, was associated in a randomized trial with a 25% reduction in the prevalence of prostate cancer documented by biopsy compared to placebo. On the other hand, the prevalence of high-grade cancer under finasteride was higher than that under placebo. By trying to understand the reasons, the authors showed that the distribution of PSA depended on the prostatic volume with a decrease in the surface under the ROC curve (see chapter IX) associated with the increase in prostatic volume, and this according to the grade of the cancer. In other words, PSAs were more effective in diagnosing high-grade cancer in small-volume prostates compared to enlarged prostates, and the apparent increase in high-grade prostate cancer in the finasteride group was at least in part, linked to their detection facilitated by the reduction in volume of the prostate… itself in relation to the taking of finasteride [12].

A detection bias can therefore, depending on the pathology studied and the study conditions, play in the direction of the diagnosis of less serious pathologies with apparent, but artefactual, improvement in the prognosis, or in the direction of the diagnosis of more serious pathologies with apparent worsening of the prognosis.

Each study can, depending on its structure and the pathology involved, generate its own selection biases and produce new ones, not yet catalogued. Their research must be systematic in the construction of the study and systematic again when reading the results.

B- Measurement biases:

Since the selection of patients (and controls, if applicable) is carried out in the least biased way possible, measurement biases, also called information biases or misclassification, may still occur: the object of study (the risk factor in a case-control study, the incidence of a disease in a prospective cohort study, the effect of a treatment in a randomized trial, the prevalence of a risk factor or disease in a horizontal study) was incorrectly measured. The reasons can be many.

The diagnosis of a disease, the measurement of an effect, the determination of a risk factor depend closely on the sensitivity and specificity of the method implemented to recognize them (diagnostic test, complementary examination, qualitative scale or quantitative, quality of the question asked during the interrogation).

Examples:

- Colorectal cancer is one of the first cancers affecting both sexes, and screening for polyps or small lesions, if possible by non-invasive means, should help prevent progression to invasive cancers. Capsule video in a recent trial was compared to conventional colonospy. For polyps larger than 6 mm in diameter, the Se (sensitivity) did not exceed 64% and the Sp (specificity) 84%; similar figures were found for advanced adenomas. The video capsule applied to large populations would lead for the moment to many false positives and false negatives, and therefore, if we wanted to compare the prognosis of a screened and unscreened population, to a bias towards zero by 'mixing ' of the two populations [13].
- The anatomo-pathological examination is usually considered as the 'gold standard', for the diagnosis of benign or malignant lesions: we have no better for the moment than the microscope, possibly helped by immuno-markers or molecular biology, for the etiological diagnosis of a lesion. Four pathologists specializing in gastrointestinal pathology were interested in their inter-observer reproducibility on the examination of colorectal polyps, to find that the reproducibility coefficient kappa, taking into account concordant diagnoses by simple chance, did not exceed… 49% [14].

It is of course necessary to surround oneself with a maximum of precautions in the use of diagnostic tests, or exposure measurements. Despite this, careful study of tests, including gold standard tests, highlights their limitations, and the results of studies using them should be interpreted with these limitations in mind. The study of reproducibility, in particular, highlights the limits of validity of a test, even when it is applied by an experienced doctor.

The recall bias to which retrospective studies are particularly exposed is another frequent driver of information bias.

Example:

- Legal abortions have been reported to be a risk factor for breast cancer. Lindefors-Harris et al. compared two case-control studies, the first using as data source interviews with patients and their controls, the second collecting information more objectively from the Swedish national register of legal abortions. The relative risk of breast cancer with respect to previous abortion was found to be 1.5 times greater in the interview-based study than in the register-based study (significant difference). Moreover, the ratio between the underestimation of

abortions in controls and the overestimation in cases (questioning data compared to objective registry data) turned out to be equal to 22.4! [15].

Sick subjects tend to want to find an explanation for their illness, while healthy controls tend to forget their medical history... Recall bias can therefore explain abnormally high odds ratios, poorly reflecting reality and leading skewed conclusions. However, the extent of the error in the result depends on the conditions of the factors examined (risk factor in a case-control study) and can be minimal when the prevalence of the risk factor is low [16].

Measurement biases can be encountered on all types of variables studied, whether quantitative variables such as biochemical assay (assay techniques have their imprecision), semi-quantitative variables such as the assessment of a tumor stage (I, II, III, or IV), because there are limits to the detectability of secondary visceral or lymph node locations, to qualitative variables such as some questioning data (personal or family history "forgotten") or anatomo-pathological examination or radiological (in favor, or not, of such and such a diagnosis).

Finally, there are **misclassification biases** linked to errors in the manipulation of data, errors in filling out questionnaires, or even false data knowingly provided by unscrupulous investigators...

When this misclassification bias occurs in a bi-directional way (group A patients inadvertently included in group B, and group B patients inadvertently included in group A), the two groups that we sought to compare, which were initially assumed to be different, become uniform. The difference then disappears artificially and can even cancel out. This is the Anglo-Saxon **"bias toward the null"**. This general truth can suffer some exceptions, in case of multiple exposure in particular [17].

When the bias is uni-directional, the result will depend on the direction of the bias! In a retrospective study with overestimation of the risk factor in the patient group and underestimation in the control group (case most often encountered), there will be an increase in the odds ratio and therefore an overestimation of the risk. In a prospective cohort study where those lost to follow-up would be found mainly in exposed subjects (example: exposed employees of a company that has gone bankrupt, forced to leave the region for professional reasons), the relative risk can be artificially lowered by one-way missing information on the lost to follow-up (actually unusual scenario in practice).

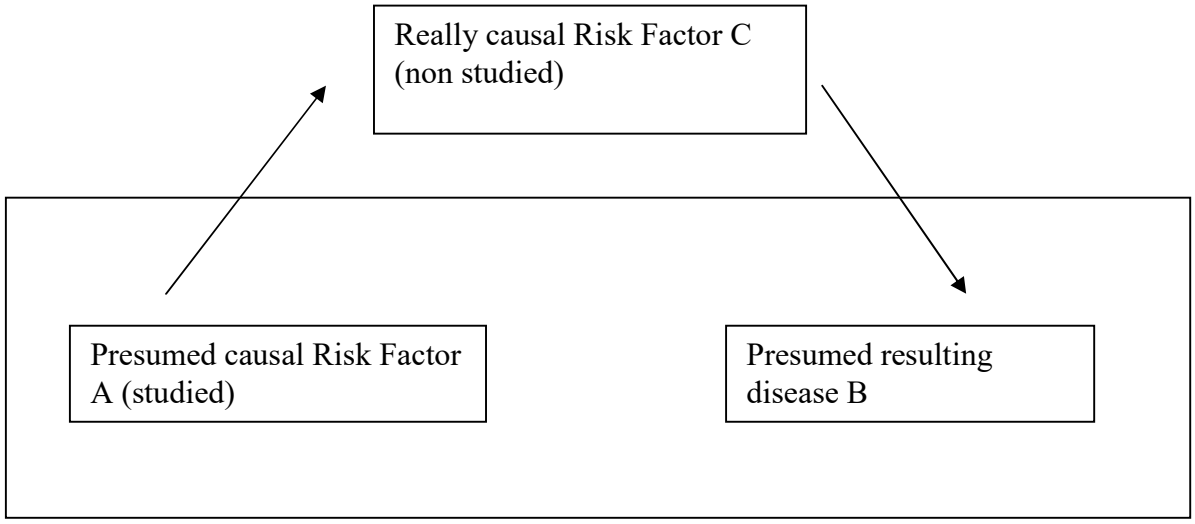
B- The confounding elements:

Independent of the first two types of bias, confounding bias occurs when the statistically observed association does not correspond to a biological reality, pathological or etio-pathogenic, but is in fact explained by a third factor, really involved in the pathophysiology of the disease studied.

Imagine a cohort study studying the effect of risk factor A on the occurrence of disease B. A cohort study cannot test many etiological hypotheses (unlike a retrospective case-control study), and the risk factor A must therefore be strongly suspected of being causal before the start of the prospective study. Now imagine that we observe, in the part of the population subject to risk factor A, a higher incidence of disease B than in unexposed subjects. To conclude that A is truly an etiological agent seems logical: the risk factor was present before

the disease declared itself, and precisely represents the distinguishing factor between the subjects who became ill and the subjects who remained healthy.

This is true on one condition: that there are not one or more C factors, or confounding elements, not measured in the study, presenting on the one hand a statistically significant association (corresponding to a biological reality or not) with the risk factor A studied, and actually involved on the other hand in the genesis of the disease B.



Factor C, which is really causal but ignored in the context of the study, therefore reveals, because of its association with factor A, the latter to be responsible for disease B in a fallacious association, statistically significant but biologically unfounded.

Immediate consequence:

• The true etiological agent will not have been recognized despite the research work undertaken.

Possible consequences:

• An action to eradicate factor A with the aim of preventing the onset of disease B may prove to be completely ineffective, provided that the association between A and C is only statistical and not biological. In other words, the elimination of factor A, if A is not physically linked to C, will not lead to the elimination of the true etiological factor and will not have an effect on the incidence of the disease.

• A therapeutic agent developed against factor A with the aim of curing disease B will not achieve the desired effect.

• Research programs can be launched on the wrong track (a non-exceptional situation that sometimes makes it possible to advance knowledge!).

Example: AIDS is a disease whose etiology is known, but transmission of the virus is more frequent in so-called at-risk groups. The high prevalence of the disease in the Bronx of New York motivated a study to determine the specific risk factors of resident drug users. The prevalence of AIDS has been found to be three times higher in black patients compared to Caucasian patients, and the temptation to conclude that there is an increased susceptibility to infection in the former is easy...

In fact, the type of drug used was not the same in the two groups for economic reasons. Cocaine was used preferentially by blacks, and heroin mainly by whites. Cocaine, because of its shorter duration of action, required three to four times more injections than heroin and the potential risk of transmission of the virus was increased accordingly [18]. An analysis of the data controlled for the confounding factor (type of drug) would have allowed a more precise conclusion.

Confounding bias can creep insidiously into the best studies, at different stages of its completion. Thus, Vach and Blettner describe how the correct use of methods for processing missing data in case-control studies can lead to the creation of a confounding bias that is independent of the construction of the study itself, but appears during the study phase. to analyze. The remedy consists in studying more precisely the distribution and the reason for the missing data in the different subgroups [19].

Studies carried out using computerized databases are particularly prone to confounding bias. Databases represent a potential source of often very rich information, but they have not been constructed with a view to specific study. They cannot therefore take into account the different possibilities of bias inherent in each study, and do not offer the possibility of remedying them.

Example: The role of beta-blockers in the prevention of coronary heart disease in hypertensive patients can be analyzed within the framework of databases that have already been compiled gathering prescriptions and drug delivery, as they exist in Scandinavian countries. Beta-blockers, however, are prescribed for different indications, including the treatment of angina pectoris, and the entanglement of pathologies (hypertension, angina, post-infarction) in a same person makes the analysis particularly difficult.

The use of information gathered for a purpose other than that of the study that one wishes to carry out (economic purpose, evaluation of drug consumption and its evolution for example) is always subject to caution, and the conclusions of the study should be interpreted with equal caution [20].

Correction of confusion bias:

It is necessary, so that the confounding element can be eliminated during the analysis of the data, that it was planned to collect the information concerning it, during the construction of the study, so that it can be controlled . The control consists of establishing subgroups of patients according to the suspected confounding factor, and calculating the association between the disease and the risk factor studied within each subgroup.

Example based on fictitious data inspired by published articles: the risk of infection by the HIV virus seems to be increased among drug addicts of an ethnic group A, compared to another (B)

	HIV +	HIV -	
Group A	287	163	450
Group B	163	287	450
	450	450	

$$OR = \frac{287x287}{163x163} = 3.1$$

If we admit that the subjects of group A preferentially used cocaine and shot themselves four times more often, the type of drug used then represents a confounding element that must be controlled: we construct ***sub-groups on the basis of the drug used***:

	Cocaine		**Heroin**		
	Group A	Group B	Group A	Group B	
HIV+	275	138	12	25	450
HIV-	25	12	138	275	450
	300	150	150	300	

Then we calculate the odds ratio for each subgroup:

1- Cocaine sub-group:
$OR = \frac{275 \times 12}{25 \times 138} = 0.95$

2- Heroin sub-group:
$OR = \frac{12 \times 275}{138 \times 25} = 0.95$

and we see that the difference observed initially disappears and was to be put down, in fact, to the drug used and the number of daily injections.

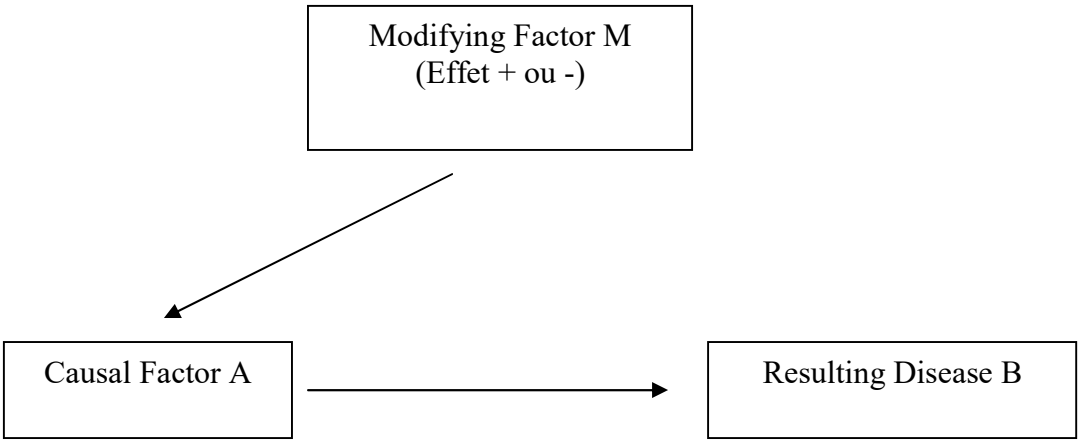
Control by the confounding element during the statistical analysis is of course only possible if the potential confounding element is already suspected during the construction of the study and if the data are collected accordingly!

Another method of statistical control of the confounding element can prove to be more powerful, because it does not require the division of the initial sample into subgroups: the ***multivariate analysis in logistic regression***, in which the potential confounder can be introduced as an independent variable. If the greatest part of the explanatory power goes to the potential confounding element to the detriment of the initial assumed risk factor, the confounding element will be retained in the model as significant, and the initial assumed risk factor will see its significance reduced, or even cancelled.

D- Effect modifiers, or modifying factors:

We have seen that a confounding factor represents a genuine causal factor, the ignorance of which in a study causes the causal responsibility to be wrongly attributed to another factor measured in the study and statistically but not biologically associated with the occurrence of the disease.

It is important to distinguish between confounding factors and modifying factors: a modifying factor is, by definition, an agent that will modify the relationship between a truly causal factor and the resulting disease. There is interaction between the causal factor and the modifying factor. Several scenarios may arise:



1- The modifying factor M does not play a pathogenic role if it is isolated, but accentuates the role of the causal factor A when A and M are present simultaneously. For example, the delta virus alone cannot cause hepatitis. In association with the hepatitis B virus, it reinforces the pathogenic role of the latter with aggravation of the hepatitis and more frequent evolution towards chronicity and cirrhosis. *M alone cannot cause disease B*.

2- The modifying factor M does not play a role when it is isolated, but attenuates the role of the causal factor A when they are associated. This is the case of genetic or acquired protective mechanisms against an illness: the sickle cell disease gene plays no role in malaria infection. However, its presence in the heterozygous state decreases the severity of the disease caused by Plasmodium by reducing the capacity of the latter to parasitize red blood cells.

3- The modifying factor M plays a role in the genesis of the disease, but its presence associated with the causal factor A results in a multiplication of effects and not in the addition of their separate effects: M and A are synergistic. Tobacco and hypertension are two well-known risk factors for cardiovascular disease. Their association, however, leads to a multiplication, and not an addition, of the risk.

4- The modifying factor M plays a role in the genesis of the disease, but its presence associated with the causal factor A decreases its pathogenicity: M and A are antagonists. This situation, unfortunately quite rare in Medicine, could illustrate the old theory of Hippocrates of the fight of evil by evil? It would thus seem that the increase in cholesterol in the brain observed in alcoholic patients would in fact antagonize the action of alcohol on neuronal connections [21].

5- The antagonistic effect can sometimes be exerted... to the detriment of a final benefit. Several studies have shown a reduced risk of colorectal cancer in patients taking calcium and vitamin D to prevent osteoporosis. This beneficial effect was not found in a randomized trial, one arm of which also included estrogen treatment. It turns out that in this trial, estrogens acted as a modifier of antagonistic effect on the action of calcium and vitamin D, to cancel the beneficial preventive effect [22].

How, in epidemiological practice, to differentiate confounding factor and effect modifying factor?

The analysis controlled by the confounding factor (the type of drug in the previous example) made it possible to eliminate the responsibility of a factor wrongly assumed to be causal (the ethnic group). The effect-modifying factor-controlled analysis will likewise establish the role of each causal factor. Imagine that a cohort study on the harmful effects of tobacco in ENT pathology gives the following results after a 10-year follow-up:

	Tobacco +	Tobacco -
ENT cancer +	200	60
ENT cancer -	600	740
	800	800

As this is a cohort study, we calculate the cumulative incidences over 10 years of ENT cancers in the group of smoking patients and the group of non-smoking patients:

Tobacco group: Incidence = 200/800 = 25%
Non-smoking group: Incidence = 60/800 = 7.5%

Since alcohol is known to be a risk factor for ENT cancers, the study coordinators had planned to measure its consumption, along with that of tobacco, among all the members of the cohort. An analysis controlled by the alcohol factor is carried out, which gives the following results:

	Alcohol +		**Alcohol -**	
	Tobacco +	Tobacco -	Tobacco +	Tobacco -
ENT cancer +	160	40	40	20
ENT cancer -	240	360	360	380
Total	400	400	400	400
Incidence	40%	10%	10%	5%

First of all, we see that controlling for the alcohol variable does not make the relationship between tobacco and cancer disappear: the incidence of ENT cancers in smoking subjects is always higher than in non-smoking subjects. Alcohol is therefore not a confounding factor in the tobacco-cancer relationship, which really exists.

We then see that the group (tobacco - alcohol -) has a risk of ENT cancer of 5%, the groups (Tobacco + Alcohol -) and (Tobacco - Alcohol +) an equivalent risk of 10%, and the group (Tobacco + Alcohol +) a risk not of 20% (sum of the separate risks of Alcohol and Tobacco), but of 40%. We are in situation 3, where alcohol alone is a risk factor for the disease (because the group without alcohol has a lower risk than the group with alcohol) and a factor modifying the toxic risk of tobacco in l increasing.

Finally, we note that the raw incidence of 25% in the group (Tobacco+) is between the incidence of 40% in the group (Tobacco+Alcohol+) and the incidence of 10% in the group (Tobacco+Alcohol-). Similarly, the crude incidence of 7.5% in the (Tobacco-) group is between the 10% incidence in the (Tobacco-Alcohol +) group and the 5% incidence of the (Tobacco-Alcohol-) group. This relationship, which can be summarized as follows, is characteristic of the role of a modifying factor of the variable M by which the control is made in the analysis:

Incidence group M - < raw incidence < Incidence group M +

When the exact role (confounding factor or modifying factor?) of a variable in the genesis of a disease is unknown, the analysis controlled by this variable makes it possible on the one hand to recognize its character, and on the other hand, in the case of a modifying factor, to quantify the direction (synergy or antagonism?) and the exact extent of the modification of the effect of a causal factor.

In our example, smoking or alcoholism taken alone multiplies the risk of ENT cancer by 2 (10% versus 5%), and, combined, by 8 (40% versus 5%).

E- how to interpret an epidemiological study ?

The structural biases of the clinical studies that we have just detailed probably explain many of the divergent results observed in the literature. However, other factors should not be overlooked:

1- The quality of the statistical analysis:

The type of statistical analysis, the tests used, the application to the questions asked **must be defined as soon as the study is constructed**. This makes it possible on the one hand to collect the data in the appropriate form for the analysis, thus avoiding inopportune manipulations, and on the other hand to rigorously test the initial hypotheses by avoiding the multiplication of calculations carried out randomly on the data. In other words, the multiplication of comparisons, of association test calculations exposes the discovery and description as statistically significant of differences or associations for which the significance coefficient is significant only by chance.

2 examples:

- Some statistical software make it possible to establish correlation matrices between all, or part, of the data collected. These are statistical correlations, not necessarily corresponding to a biological or epidemiological reality. If we accept an alpha error coefficient of 5% and if we perform 100 correlation tests at random, 5 of these tests may prove to be falsely significant because they correspond to the 5% error allowed. There is a great risk then, if only the 5 positive results are published without indicating that they were observed by chance in a batch of 100 tests, of bringing to the attention of readers only a statistical artefact. This is the publication of results that do not meet a pre-established research hypothesis.
- All software makes statistical analysis very easy and fast, compared to the time required for calculations carried out manually… And therefore, they all expose to the risk of multiple tests without a pre-established rationale, and to the risk of significant p values just … by chance.

The commonly accepted 5% alpha error is an arbitrary choice. The risk at the end of a well-conducted statistical analysis of obtaining a positive result by chance or chance always remains, and it is sometimes enough to add a few additional cases to a series for alpha to increase to 6 or 7% and that the results are held as insignificant... Lowering alpha increases the reliability of the results.

2- Unpublished studies:

A study with negative results generally has less chance of being published than a study with positive results (aside from large multicenter studies). There is therefore a bias in the literature **towards positive results**, which is particularly important from a practical point of view for therapeutic studies when it would be important to know that the effectiveness of a treatment is not found regularly in different investigative centers. **The publication bias** is an invisible bias if there is no registry of all the studies carried out, published or not. When such a registry exists, comparing the published results with the results actually obtained, but sometimes not published, highlights a significant difference and a clear influence of non-publication bias [23].

3- Any study must be interpreted in the broader context of the knowledge already acquired:

In 1965, B. A. Hill proposed **9 criteria for interpreting** an epidemiological study [24]:

1- **Strength of the observed association**: an association is all the more likely to be real as its measured amplitude is greater. Relative risks below 2 are low.

2- **Reproducibility**: a result will be all the more reliable if it is found, as in biology, by different investigators.

3- Specificity: a specific causal relationship between a risk factor and a given disease is an additional argument for admitting the responsibility of the factor. Example: BK and tuberculosis, whereas one will hesitate to recognize as pathogenic of a pulmonary lesion germs found in the sputum... A risk factor can however intervene in different diseases.

4- **Temporality:** establishing that the supposed cause precedes the observed effect represents a strong argument!

5- **Biological gradient:** the increase or decrease in the incidence of a disease under the influence of the increase or decrease in the frequency or intensity of exposure, makes the pathogenic responsibility of the exposure more likely.

6- **Plausibility:** its definition obviously suffers from many biases, but common sense remains useful even in the most advanced scientific approaches!

7- **Consistency:** results will be more likely if they fit within the framework of knowledge acquired as solid without upsetting them too much... There are of course notable historical exceptions.

8- **Experimentation:** this is not always possible in Epidemiology, but can come to consolidate data observed in the field.

9- **Analogy:** this criterion also has its limits, but the resemblance of an observed phenomenon to other already known phenomena can increase its credibility.

4- Application of the results of a study to an individual:

Group studies are necessary in order to better understand the overall reality of a pathology or a treatment, which is impossible to grasp from just one or a few patients. They make it possible to observe general trends at the group level, and the conclusions will be applied at the group level with reasonably foreseeable results *because they are expected on the data collected at the group level:*

Example: Vaccination against smallpox has eradicated the disease from the surface of the globe, thus reducing the incidence to 0/year/million inhabitants.

Corollary: individuals seem currently definitively protected against the disease at the cost, however, of a certain number of serious vaccinal encephalitis, occurring in patients who perhaps would never have been affected by the disease: *the individual benefit can be dissociated from the group benefit*.

In practice, the clinician is often confronted with situations where the respective weight of benefit and risk for a specific individual is much more difficult to assess, despite the performance of extensive studies:

Example: Most studies of thrombolysis in the acute phase of myocardial infarction have shown a reduction in hospital mortality in treated patients. These are therefore confirmed results, and all the more reliable as the magnitude of the reduction is similar from one study to another (approximately 25%). In the GISSI study, overall mortality fell from 13% in the untreated group to 10.7% in the treated group (reduction of 2.3% in absolute value, from 18% = [(10.3 - 13)/13] x 100 in relative value) [25].

However, neurological accidents under thrombolysis do not exceptionally occur, in 0.94 to 1.33% of cases [26]. Treating 1000 patients with thrombolysis in the acute phase amounts to avoiding 23 deaths during the period of hospitalization, but to causing 10 to 13 presumed hemorrhagic neurological accidents. The benefit is greater than the risk and the treatment is therefore legal, but the clinician facing his patient has very few means of knowing, even respecting the contraindications of thrombolytics, if:

1- his patient is one of the 23 patients at risk of death if treatment is not administered;
2- his patient is not one of the 13 patients at risk of cerebral hemorrhage under treatment;
3- his patient is not one of the 977 patients for whom the benefit of treatment in terms of survival is nil.

The epidemiologist can seek to better define the prognostic factors, the indications for treatment, the characteristics of unimproved patients. However, its methodological tool is not adapted to the particular study of the individual and it can only provide a general (valuable) orientation to be modulated by the clinician according to the present case. Decision trees seek to fill this gap by integrating the particularities of each patient, but they remain difficult to apply and cannot be used routinely.

CONCLUSION

Epidemiological techniques applied to pure epidemiological research or to clinical research therefore have, like any technique, their limitations both in their field of exploration and in their possibilities of interpretation. They nevertheless remain essential and unequaled until now each time one wants to measure a phenomenon affecting living beings in their normal environment, with all the interactions characterizing them, outside of artificial conditions of isolation from a laboratory. They remain the only tool capable of verifying in the field the hypotheses generated in vitro or by experimentation. Finally, they make it possible to attribute to each factor the weight which is due to it in reality in its pathogenic, physiopathological or therapeutic action... but they must, because of the complexity of the phenomena they cover and attempt to describe in a global manner, to be handled with knowledge of their faults. They

are to medicine what music theory is to music: necessary, but insufficient to interpret the score and adapt it to the present moment.

References

1 Greenfield ML, Mhyre JM, Mashour GA, Blum JM, Yen EC, Rosenberg AL. Improvement in the quality of randomized controlled trials among general anesthesiology journals 2000 to 2006: a 6-year follow-up. Anesth Analg 2009;108:1916-1921.

2 Brenner H, Savitz DA. The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control studies. AMJ Epid 1990;132:181-192.

3 Holt VL, Daling JR, Stergachis A, Voigt LF, Weiss NS. Results and effect of refusal recontact in a case-control study of ectopic pregnancy. Epidemiology 1991;2:375-379.

4 Doll R, Peto R, Hall E, Wheatley K, Gray R. Mortality in relation to consumption of alcohol: 13 years' observations on male British doctors. BMJ 1994; 309: 911–918.

5 Austin PC, Mamdani MM, van Walraven C, Tu JV. Quantifying the impact of survivor treatment bias in observational studies. J Eval Clin Pract 2006;12:601-612.

6 Di Martino LD, Hammill BG, Curtis LH, Gottdiener JS, Manolio TA, Powe NR, Schulman KA. External validity of the cardiovascular health study: a comparison with the Medicare population. Med Care 2009;47:916-923.

7 Ioannidis JP, Polycarpou A, Ntais C, Pavlidis N. Randomized trials comparing chemotherapy regimens for advanced non-small cell lung cancer: biases and evolution over time. Eur J Cancer 2003;39:2278-2287.

8 Robins J, Pike M. The validity of case-control studies with non-random selection of controls. Epidemiology 1990;1:273-284.

9 Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. NEJM 1978; 299:1089-1094.

10 Hulka B, Grimson RC, Greenberg BG, Kaufman DG, Fowler WC, Hogue CJR, Berger GS, Pulliam CC. "Alternative" controls in a case-control study of endometrial cancer and exogenous estrogen. Am J Epidemiol 1980;112:376-387.

11 Sagawa M, Usuda K, Aikawa H, Machida Y, Tanaka M, Ueno M, Sakuma T. Lung cancer screening and its efficacy. Gen Thorac Cardiovascular Surg 2009;57:519-27.

12 Elliott CS, Shinghal R, Presti JC Jr. The influence of prostate volume on prostate-specific antigen performance: implications for the prostate cancer prevention trial outcomes. Clin Cancer Res 2009;15:4694-4699.

13 Van Gossum A, Munoz-Navas M, Fernandez-Urien I, Carretero C, Gay G, Delvaux M, et al. Capsule endoscopy versus colonoscopy for the detection of polyps and cancer. NEJM 2009;361:264-270.

14 Wong NA, Hunt LP, Novelli MR, Shepherd NA, Warren BF. Observe agreement in the diagnosis of serrated polyps of the large bowel. Histopathology 2009;55:63-66.

15 Lindefors-Harris BM, Eklund G, Adami HO, Meirik O. Response bias in a case-control study: analysis utilizing comparative data concerning legal abortions from two independent Swedish studies. Am J Epid 1991;134:1003-1008.

16 Drews CD, Greeland S. The impact of differential recal on the results of case-control studies. Int J Epid 1990;19:1107-1112.

17 Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epid 1990;132:746-748.

18 Schoenbaum EE, Hartel D, Selwyn PA, Klein RS, Davenny K, Rogers M, Feiner C, Friedland G. Risk factors for human immunodeficiency in intravenous drug users. NEJM 1989; 321:874-879.

19 Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad-hoc methods of correcting for missing values for confounding variables. Am J Epid 1991;134:895-907.

20 Psaty BM, Koepsell TD, Siscovick D, Wahl P, Logerfo JP, Inui TS, Wagner EH. An approach to several problems in using large databases for population-based case-control studies of the therapeutic efficacy and safety of anti-hypertensive medicines. Statistics in Medicine 1991, 10, 653-662.

21 Crowley JJ, Treistman SN, Dopico AM. Cholesterol antagonizes ethanol potentiation of human brain BKCa channels reconstituted into phospholipid bilayers. Mol Pharmacol 2003;64:365-372.

22 Ding EL, Mehta S, Fawzi WW, Giovannucci EL. Interaction of estrogen therapy with calcium and vitamin D supplementation on colorectal cancer risk: reanalysis of Women's Health Initiative randomized trial. Int J Cancer 2008;122:1690-1604.

23 Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009;302:977-984.

24 Hill BA. The environment and disease : association or causation ? Proc Royal Soc Med 1965;58:295-300.

25 Gruppo Italians per studio della Streptochinasi nell' infarto miocardio (GISSI): Effectiveness of intraveinous thrombolytic treatment in acte myocardial infarction. Lancet 1986;1:397-402.

26 Gruppo Italiano per lo Studio della Sopravinenza nell' Infarto miocardico : GISSI 2 : a factorial randomized trial of alteplase versus streptokinase and heparin versus non heparin among 12 490 patients with acute myocardial infarction. Lancet 1990;336:65-71.