# Case-control studies in psychiatry: causalty, design and warnings

L. C. Silva Ayçaguer<sup>a</sup>

<sup>a</sup> Instituto Superior de Ciencias Médicas de La Habana. Cuba

#### Estudios de casos y controles en psiquiatría: causalidad, diseño y advertencias

#### Summary

This present paper is mainly methodological and has been written with the aim of helping researchers in psychiatry to produce results with higher quality and help readers to have adequate assessment values of others. Brief reflection is made on the most important conditions that must be fulfilled to prove a causality bypothesis, regardless of the investigation design used. However, the main purpose of the text is to examine and illustrate how these conditions work under a case-control study environment. Besides outlining the basic aspects concerning design and analysis, areas extremely illustrated with examples of case and controls in psychiatry found in the literature, a number of suggestions to avoid pitfalls that can invalidate research efforts developed using case-control methodology is offered.

Key words: Case-control studies. Biases. Stepwise regression. Prediction. Logistic regression. Sample size.

#### Resumen

El presente trabajo es de naturaleza esencialmente metodológica y se ha escrito con la expectativa de ayudar a que la investigación en psiquiatría se lleve a delante con creciente calidad, así como para contribuir a que los lectores de trabajos ajenos dispongan de pautas valorativas adecuadas. Se reflexiona brevemente en torno a las condiciones más importantes que necesariamente han de cumplirse para probar una bipótesis de causalidad, independientemente del diseño de investigación de que se trate. Sin embargo, el núcleo central del texto se destina a examinar e ilustrar el problema que tales condiciones plantean a quien diseñe y conduzca un estudio de casos y controles, que constituye la expresión dominante en la investigación observacional contemporánea. Además de reseñar los aspectos básicos en materia de diseño y análisis, áreas profusamente ilustradas con ejemplos de la literatura actual sobre casos y controles en psiquiatría, se ofrece un conjunto de advertencias sobre aspectos que suelen no contemplarse y cuyo olvido deteriora y ocasionalmente invalida las investigaciones desarrolladas con este importante método.

Palabras clave: Estudios de casos y controles. Sesgos. Regresión paso a paso. Predicción. Regresión logística. Tamaño muestral.

# 1. INTRODUCTION

In general, clinical and epidemiological investigation is undertaken either to describe a reality or condition (characterize how it is in relatively few features) or to explain why this reality is the way it is, which is essentially equivalent to discovering or corroborating the causal mechanisms that govern it.

Several conditions must necessarily be fulfilled to «have the right» to believe that a certain condition casually

Correspondence:

Luis Carlos Silva Ayçaguer Instituto Superior de Ciencias Médicas de La Habana Vicerrectoría de Investigaciones Calle G y 25, 6.º piso, plaza Ciudad de la Habana (Cuba) E-mail: lcsilva@infomed.sld.cu influences in another. To mention a source where they are listed, the important list of conditions of Hill<sup>1</sup>, which has been complemented with other demands<sup>2</sup>, is generally cited. However, not all these demands have the same hierarchy. For example, a main premise is that the casual hypothesis is plausible. This is a requirement prior to any action that is methodological and independent of it and it constitutes the first unavoidable sieve. Besides the fact that it may have a certain empirical backing, this means that it should have a rational theoretical support, and, in any event, it cannot have a direct contradiction with facts established by science. This demand is far from being a lateral adornment. Absence (or vague formulation) of a relatively persuasive basis that confers plausibility to the hypothesis is not an uncommon phenomenon. Unfortunately, studies conceived from more or less spectacular anecdotes, many times related with undefined energies that no one has been able to measure, are not rare.

An extreme example, but perhaps for this same reason, eloquent, is offered, in my opinion, by homeopathy<sup>3</sup>, a therapeutic modality that has awaken some interest within the framework of psychiatry. In an article in which the role of this therapeutic resource in the treatment of social phobia is exalted, that has been published in a recently founded Journal to publish studies on the so-called alternative therapies, Davidson and Gaylord<sup>4</sup> admit that «there is no acceptable explanation for homeopathy within the framework of the present medical theory». They also remind us of a doctrine that is proclaimed from its own title in another article of this same journal (The mechanism of homeopathy: all that matters is that it works)<sup>5</sup>. I consider that, unless financing is abundant or there are other social circumstances advising it, there is no greater sense in examining the therapeutic resources which, as this, not only lack all foundation, but are also difficult to reach at some time, since their followers (perhaps knowing that they would not find it) boast of not needing it.

In the present paper, however, emphasis will be placed on the examination of those causality premises that are directly related with methodological aspects, especially on those which, in my opinion, are truly crucial or unavoidable - that is, existence of association, temporal congruence, absence of biases responsible for the observed association - and that the latter cannot be explained by the action of other factors (the so-called confounding factors). Whatever the design used, it must be compared with the four conditions mentioned above. Section 2 of this paper is aimed at briefly commenting on each one of them.

In any area of medical knowledge, and therefore in the field of psychiatry, the methodological resources in design material to corroborate etiological conjectures or, more generally, causal hypotheses, cover a wide range of variants and subvariants. However, mentioned concisely and essentially, these procedures are basically three: case and control studies, longitudinal or cohort studies and experimental ones. The latter, of which the controlled clinical trial is the most emblematic expression in the biomedical world, are the optimum way of approaching the causality problem. The prospective character inherent to all experiment solves the possible obstacle of temporality and, if the clinical trial is adequately directed (especially when making techniques and random allocation are used), many of the possible biases and the effect of third factors are conflicts that are cast out by the design itself.

Unfortunately, ethical and practical imperatives having diverse nature generally make experimentation impossible. The observational cohort studies are the natural substitute for the clinical trial, since they share its prospective nature. The facts are recorded in time ascending order, observation begins both for those who have a presumed risk condition as well as those who do not have it, when the outcomes studied have not yet been expressed. In essence, an attempt is made to compare the incidence rates between two groups, so that, except for rare situations (see article of Silva and Benavides<sup>6</sup>), adequate recording of temporality is guaranteed beforehand. But this approach lengthens the lapses to reach analyzable results (sometimes for years) and is generally prohibitively expensive.

Such circumstances made the case and control studies the adequate resource when resources are limited and available time reduced; and this may be why they have become the dominant form of contemporary etiological investigation. This approach, however, is not a panacea: it can only produce high quality results when the different methodological observations characteristic of it are overcome. These results are even similar to that of those generated by behavior of the clinical trials (see the series of papers recently published in Lancet by Schulz and Grimes, especially that which is dedicated to this type of studies<sup>7</sup>). The basic traits that define this procedure are explained in section 3. I try to reflect on the most relevant aspects in regards to design as well as to outline the fundamental analyses lines that correspond to it. However, since there is abundant specialized bibliography on the subject, including books<sup>8,9</sup> and entire issues of journals (for example, a monographic number of Epidemiologic reviews\*), the treatment of the subject in this section points to the management of the most general aspects and is developed stressing conceptual more than operative aspects.

Section 4, finally, concentrates on sharing some warnings that, although they concern errors that are relatively easy to avoid, must be taken into account to avoid consequences that may be catastrophic, as will be duly illustrated. Apart from that, it deals with warnings that are insufficiently explained in a reasoned and understandable way in the non-specialized literature.

Our hope not only is to give guidelines or ideas that tend towards increasing the excellency of investigation in psychiatry, but that also contribute to the development of the scientific article readers' critical capacity, especially in regards to the many papers in which the results obtained with case and control methodology are given.

# 2. MAIN CONDITIONS FOR CAUSALITY

# 2a. Association

The existence of associations is a clear premise of causality. However, in practice, and in spite of repeated warnings, it is continuously confused with the causality itself. It is not rare to find projects, dissertations and even articles published in prestigious journals whose objectives are to identify the association existing between two variables. The formulae used for this declaration vary: in some cases, the purpose of «evaluating if there is correlation» is announced. Others report that they want to

<sup>\*</sup>It is number 1 of volume 16, published in 1994.

«establish the existing association», and some directly establish that they want to «correlate» a variable (typically, a presumably causal factor) with another (in general an «outcome» of a certain process). Unfortunately, considered in this way, it lacks sense. The act of calculating an association coefficient does not mean more than the mechanical application of a method, and therefore cannot be a genuine objective of investigation. The fear to frankly and clearly admit that they want to «prove that X influences Y» or «to assess the degree in which X is the cause of Y» is very extended and it is decided to use the subterfuge of stating that they only want to «quantify the association between X and Y», as if the association could have an intrinsic interest. Simply, the objective is substituted with one of the means to obtain it.

#### 2b. Temporal congruence

We all know that a causal link means a temporal relationship in the sense that the presumed cause of a certain effect, necessarily, must precede it in time. Since the time when we were very young children, we have understood that light in a room is associated with the fact that the switch must be used. However, due to the order in which these events are inexorably produced, we also quickly learn that this is the final action that produces light, and not vice versa. No causality analysis has a clear meaning when the study design has not included the precaution of recording the events in such a way that a rule that is so basic as the statement can be contemplated. The danger of not respecting it is very important in the case-control studies. In Section 3b, it can be seen that it occurs with a far from negligible frequency.

#### 2c. Effect of confounding factors

A possible explanation of the fact of having observed an association between two variables is that it is a mere reflection of the action of a third factor on both; in such case, our observation should be attributed to a relationship having a structural nature in spite of dealing with something that is simply phenomenological. In order to avoid this interpretative error in non-experimental studies such as those of cases and controls, special techniques, such as those outlined in section 3c, must be used.

#### 2d. Biases

It has been stated that performing a good investigation is reduced to avoiding the biases found in all design. Such is the case, to give an example, of the so-called interrogator bias, consisting in the fact that the person who questions or measures the participants, for whatever reason, does not place the same zeal in obtaining information for all the elements (he/she gives less care when recording data for the healthy subjects than for the patients, or is more inquisitive if it is a young subject rather than an elderly one). The inventory of possible biases is great (from anticipation in the diagnosis, of the interrogator, of memory, of selection, etc.) and the literature approaching it is equally abundant. To access a detailed and profoundly illustrated description, I recommend a recent and careful book of Moises Szklo and Javier Nieto<sup>10</sup>. In section 3d, the two most important biases that are generally found in case-control studies are dealt with.

#### 3. DESIGN OF CASES AND CONTROLS

Having made this general review, and having explained the four basic premises to convalidate a causality hypothesis in this context, we consider the specific problems that establish the management of each one of them for those who operate with a case-control study in this section.

The fundamental trait of this methodology consists in attempting to solve the causality problem by a reconstruction process of the events, at the end of which two groups have been compared: that of the individuals who have had a specific outcome and that of those who have gone though a similar process, but have resulted in an opposite outcome. The frequency with which the factors that are causally suspected related with this outcome are recorded constitutes the object of the comparison between the two groups

Stated in another way, in these studies, a group of individuals - the cases - carriers of a given condition (usually a disease, but sometimes, another type of adverse event, as, for example, a suicide attempt) are chosen to be compared with a group of subjects - the controlswho do not show this condition. For both, past data are recorded (frequently called «exposition factors») that are considered relevant to the appearance of the morbid condition studied, such as could be birth weight, drug consumption or certain sexual practice.

These studies owe their noticeably popularity basically to the agility and economy with which they develop. They are the natural alternative to the cohort studies, that are rarer, of longer duration and more complex from the logistic point of view. During the first half of the XX century, the fundamental evolution of the method was verified, above all incidental to the desire of elucidating the etiology of the chronic diseases in a highly industrialized society, whose morbidity and mortality pattern no longer responds to the framework of the infections. The main part of the statistical methods, with which the data from this type of studies has been managed, arises in this context. One of the most important contributions is owed to Cornfield<sup>11</sup>, who suggested using the odds rations to estimate relative risk of a casecontrol investigation. In the United Kingdom, the greatest promoter of this procedure was Bradford Hill, who obtained fame from the celebrated study performed in 1950 with his student Richard Doll<sup>12</sup>. This study assessed the effect of the smoking habit on lung cancer.

#### 3a. Association

In coherence with the fact that it makes no sense to propose the identification of correlations themselves, no one considers the idea of investigating the association between the fact that an adolescent attempts suicide and, for example, the color of the room in which he sleeps. Why? Simply because, although the association between the two conditions is a condition *sine qua non* to conceive that they are causally linked, as long as it is attempted to state its existence is due to the fact that there is a suspicion (at a minimum, subconscious) that such corroboration could provide a test (or at least a sign) of a causal relationship<sup>13</sup>.

For example, when Zornberg and Jick study the relationship between neuroleptic consumption and appearance of idiopathic venous thromboembolism<sup>14</sup>; they do not limit themselves to calculating a measure of association (let's say, an odds ratio) and to performing a hypothesis test to verify that this association is not exclusively attributable to chance. In fact, what is questioned is if this consumption increases the risk of the disease, so that a combination of methodological precautions are adopted that make it possible to go beyond the mere phenomenological confirmation that the variables show association.

Although the purpose of this paper is basically conceptual, it is impossible to not introduce, although briefly, some basic mathematical elements. In the setting of the case-control studies, the procedure par excellence to measure association is, undoubtedly, the *odds ratio*.

To illustrate it, we suppose that a combination of 181 adolescents who made a suicide attempt was studied and that three adolescents of the same age and gender, chosen from the same community, who had never made an attempt against their life, were studied for each one of them. We imagine that the percentage of those who had a significant weight loss during the previous semester to the date of the suicidal act reached 70% among the former and only 56% among the latter. We have observed a clear association between suicide attempt and weight loss. The data mentioned can be summarized as seen in table 1.

The OR is calculated or estimated by the so-called «crossed products ratio»:

OR = 
$$\frac{\text{ad}}{\text{bc}} = \frac{(128)(241)}{(53)(302)} = 1.9$$

In this example, in principle, it can be considered that it is almost two times more likely for a boy who has lost weight to make a suicide attempt than one who has maintained normal weight<sup>\*</sup>. Observe in the formula that

TABLE 1.	Distribution of cases and controls according
	to whether there was weight loss or not
	before the suicide attempt

Suicida attempt	Weigh	Total	
Suicide dilempi	Yes	No	101111
Yes (cases) No (controls)	a = 128 c = 302	b = 53 d = 241 294	181 543 724

the OR numerator will be large when a and d are and the denominator small, when c and d are small. Consequently, OR tends to be greater in the measure that the data tend to be located in the first diagonal of the table, thus indicating greater association between the presumed risk factor (weight loss) and the outcome (suicidal attempt). To assess if this number is «significantly» greater than the unit, until relatively recently, a significance test was usually carried out that gave rise to a Chi squared value and a p associated value. According to the enthroned rules during decades, if the p value is less than 0.05, then it is considered that there is «significance». This is easy to corroborate by means of any of the computer programs available to the effect that, if we decide for this approach, we would obtain Chi squared = 12.8 and p = 0.0003. In general, this procedure, but especially in the OR analysis, has been becoming clearly obsolete (see section 4d). From the appearance of a convincing article by Gardner and Altman15 in the British Medical Journal, the use of the hypothesis test has been replaced with the report of the confidence limits for the OR. As before, it is very easy to presently find computer resources that make it possible to make this calculation. In the example, for this OR we see that the interval (at 95% confidence) is (1.3-2.8), which leads us to verify that we can be highly confident that the OR is, at a minimum, 2.3, although it could reach 2.8. That is, the probability of a suicide attempt is, at least, 30 % greater for adolescents who have had a noticeable weight loss in the last semester than for one who has not fulfilled this case.

#### 3b. Temporal congruence

Not taking the premise of temporal precedence into account is an open trap, especially insidious in cross-sectional and retrospective studies, in which investigation should be made on facts occurring prior to the moment of the study. The key to the problem is found in the fact that it is impossible to establish what was the order in which the facts recorded occurred by observation in these cases. The case-control studies, in which it can only be attempted to reconstruct the events, are highly vulnerable to the appearance of this problem.

In the retrospective studies, besides healthy subjects, individuals who, at the time of the survey, suffer a di-

<sup>\*</sup>Typically, the OR can be considered equal to the relative risk, a parameter that only can be estimated in the cohort studies (except it there is additional information).

sease or have a given condition are included. For example, the cases could be women who are suffering from anorexia nervosa and the controls healthy women in this sense. We suppose that for each one of them (cases and controls), data are obtained, for example family background of mental disorders, if amphetamines are consumed and if exercise is carried out. The background of a mother or father with psychiatric disorders may be a fact that is probably *prior* to the situation that this woman presents now (the paternal condition could have been diagnosed even before the birth of the child). However, to assess the possible causal effect, for example, of amphetamine consumption or of sedentary life in the development of the disease, what should really be recorded is not if the subject consumes these drugs or practices exercises at the time of the survey or shortly before, but if this was done or not during a period prior to the appearance of the condition that is sufficiently long so as to have had the opportunity to produce it.

Trying to assess if certain risk factors are specific for anorexia nervosa or more generally valid for other general psychiatric disorders or eating disorders, 67 women with the diagnosis of anorexia nervosa and 102 controls with other psychiatric disorders, as well as 102 others with bulimia nervosa were recently studied<sup>16</sup>. The investigators studied a wide spectrum of risk factors, such as obesity of the mother, mental disorders of the parents, early menarche and perfectionist personality. Observe that all these examples refer to traits that can reasonably be anterior to the possible development of the disease. Consequently, in this study, the possible associations obtained between these factors and anorexia exceed this demand, characteristic of a causal interpretation. It must be observed that there are frequently phenomena that have a mutual «feedback» as cause and effect in the socioepidemiological studies. For example, depression may be a risk factor for the development of anorexia; but this is not an obstacle for the inverse relationship to also occur: that, as a consequence of the anorexia, the patient develops a depressive syndrome or its presence becomes worse\*. I want to state that if the investigators would have included depression as a risk factor, they would have had to assure that the evaluation of such condition would have occurred before the onset of the anorexia.

When this «detail» is not taken into account, the temporal logic of the study is lost and all potential interpretation of its results is cancelled, a fact that may be a disaster, although many investigators do not take it into account, or they consider it as a lesser evil that it is «solved» by mentioning it as a study limitation.

In a very recent article<sup>17</sup>, investigation was made on the use of drugs, alcohol and tobacco by the patients with a diagnosis of schizophrenia in Scotland. For this, the consumption of such substances by 316 schizophrenics and 250 subjects obtained from the general population was determined. As a conclusion, the authors consider that «more patients reported the use of drugs the previous year (7% versus 2%) and greater consumption of alcohol (17 % versus 10 %)». Since, obviously, the disease did not begin on the day of the examination but probably several years before, these percentages would be more drug-addiction and alcoholism incidence rates between patients and healthy subjects. Unfortunately, we cannot even be sure of this, since the habit could have been established in some of the consumers before the appearance of the disease and could have even favored it. Confusing things somewhat more, the authors add: «The present smokers were 65% of the patients versus 40% in the population», and then conclude that «the problem of drug and alcohol use is greater among schizophrenics than in the general population». It is surprising to find these results in an article whose title («Use of drugs, alcohol and tobacco by schizophrenics: casecontrol study») directly refers to a retrospective study, although it reports the opposite at the end (appearance of factors among the patients).

We examine another example in some detail. In an article titled Depression and dementia: case-control study<sup>\*\*</sup>, the authors verify that they aim to know the risk factors of depression in patients with cognitive deterioration<sup>18</sup>. Their conclusions, however, are the following: «cognitive deterioration, psychiatric backgrounds and risk factors or cerebrovascular disease influence the presence of depression in geriatric samples. Gender, age, civil status, subtype, seriousness and dementia duration are not associated to the depression diagnosis.» Unfortunately, I observe several of the problems which, in my opinion, must be avoided. In the beginning, I identify the following:

- It is stated that it is a study of patients with cognitive deterioration and then it is reported that the cognitive deterioration is a risk factor.
- For some of the variables, it is «concluded» that they are not associated with depression (in place of assessing if they are risk factors or not, as announced in the declared objectives).
- It is not at all clear what are the «risk factors» that influence the presence of depression.
- The conclusion mentions «samples» which lacks meaning, since the questions are never really on the samples but rather are related to the populations that they represent<sup>13</sup>.
- The management of time is very confusing and, in any case, not productive: the authors study the causal relationships between dementia and depression, but, as they report, it is only known when the disease begins for 70% of the dementia patients while the duration ranges over a spec-

<sup>\*</sup> Such bidirectionality may be present in dementia-alcoholism, drugaddiction-sexual dysfunction, scholastic failure-social phobia, etc., binomial.

<sup>\*\*</sup> Incidentally, attention is called to the reader on the fact that I have mentioned the expression «estudio caso-control» within the text, that I consider a mimetic translation of the English «case-control study».

trum in the rest that goes from less than one year (25%) to more than 6 years (10%), and others that occurred with the depression duration (inferior to 1 year in 41%, superior to 5 years in 22% and unknown in 7%).

The dangers of not observing this crucial methodological demand is greater when chronic diseases are studied: knowledge of what occurred before its onset may be very difficult (or impossible), simply due to the difficulty (or impossibility) to identify the moment when the disorder began. In the examination of acute or specific problems, however, as occurs with suicide, all is reduced to adopting due precautions. For example Kresnow<sup>19</sup> investigates on expositions such as alcoholism, hopelessness, indirect incitements to suicide (due to behaviors of the communication media, friends or parents) and depression; the data are recorded, depending on what they deal with, for the year, month or day prior to the suicide attempt when dealing with the cases, and for the sample lapses, but prior to the moment of the questioning for the controls.

To conclude with these considerations on temporality, the important confusion that generally occurs with the concept of cases and controls can be stated. One very recent example, taken from a prestigious journal, manifests it<sup>20</sup>. Sexual dysfunctions rates (variable that is investigated by a self-administered questionnaire) are studied in 135 persons with schizophrenia and 114 subjects chosen from the general population. They obtained results such as that, among the patients, there was at least one sexual dysfunction for 82% of the men and for 96% of the women, the patients presented erection problems in a greater percentage and the women enjoyed it less than the healthy women. They conclude that «persons with schizophrenia report much higher rates of sexual dysfunction than those from the general population». Even though the title itself announces that it has performed a «case-control study», it seems that what it records, both in healthy individuals as well as in subjects that already have the disease, is the real presence or absence of dysfunctions. Such confusion is probably partially due to the amphibology of the term controls that has been generically used to refer to the subjects they are compared to. However, it is obvious that it is not legitimate to characterize the study with the denomination that is universally reserved for retrospective studies.

## 3c. Confounding factors

We suppose that in a case-control study, a certain value of OR is observed; in the beginning, this allows us to speak about whether there is an association or not between the dichotomic variables involved and, more generally, about its intensity. However, the magnitude observed could be explained (totally or partially) by the effect of a third factor that is concomitant with both; in such case, it is said that this latter is a confounding factor.

For a variable to be considered a confounding factor of the association between an exposition (or risk factor) and response (or outcome), it must formally comply with the following: be simultaneously associated with the outcome and the exposition, but without being a consequence of the latter<sup>21</sup>; that is, the possible confounding factor, to be so, cannot be a causal intermediate step between exposition and outcome (fig. 1).

The effect of a confounding factor may be both that of increasing as well as moderating the evaluation that we have on the intrinsic or structural association that really links the study variables. This is the reason that advises handing the data in such a way that, as far as possible, we can know the magnitude of the degree of the intrinsic association that these variables have. When the process that it permits is carried out, it is generally said that the confounding factor «has been controlled» or that it has obtained an estimation of the OR «adjusted» by this factor.

In the example summarized in table 1, the association observed could be explained by the fact that the presence of the depressive syndrome in the previous semester is much more noticeable among those who, when all is said and done, attempted suicide than among those who did not, combined with the fact that the depression rate is also greater among those who lost weight than among those who maintained their normal weight or gained weight. In such case, it is said that the association between weight loss and suicide attempt is confounded by the depression.

This means that the possible risk condition that was attributed to the fact that the adolescent would have lost weight is doubtful, since it is suspected that the appearance of depression symptoms could be responsible for the association observed. That is, it is not ruled out that, if this association is examined once the depression effect is «controlled», this could disappear, thus revealing that the weight loss would not be, by itself, a condition influencing the tendency to suicide, although the association observed would suggest, in the beginning, a causal link.

We suppose that when examining the data of table 1, «crossing», on the one hand, the case-control condition



Figure 1. Relationship between cause, effect and confounding factor

	Depre	Depressed		
	Yes	No	10101	
Cases	144	37	181	
Controls	211	332	543	
Total	355	369	724	
WL	417	13	430	
No WL	59	235	294	
Total	476	248	724	

TABLE 2.	Crossing of variables suicide attempt and
	weight loss with the fact of having
	had symptoms of depression or not

WL: weight loss.

with the depressive syndrome, and on the other, that of the latter condition with weight loss, we obtain the results shown in table 2.

In this relationship, the presence of a depressive syndrome complies with the requirements to be considered a variable that confounds the relationship that links weight loss with suicide attempt. In the first place, it is well known that suicide risk is greater among those who are depressed (besides the fact that the data of the first crossing endorse it, since

$$OR = \frac{(144)(332)}{(211)(37)} = 6.12$$

is a much larger number than the unit). On the other hand, depression is strongly associated with weight loss: only 5% of the non-depressed lost weight while this condition occurred in 88% of those who presented the syndrome. And finally, it does not make more sense to suppose that the supposed effect of depression occurs through weight loss (fig. 2).

In an extremely well known study, titled «Statistical aspects of the analysis of data from retrospective studies of disease», Mantel and Haenszel<sup>22</sup> introduced the resource that reigned for many years to confront the task of «controlling» the confounding factors: stratification. The key idea consists in, in fact, that when we are faced with a possible confounder, this should and may occasionally «be controlled». This means that a maneuver may be made in the analysis that makes it possible to examine the «pure effect» of the presumed cause (as if all the subjects of the sample were equal in regards to the confounding factor). The result of this maneuver is the *odds ratio* of Mantel Haenszel (ORMH), that is no more than an average of the OR observed in the strata\*.

To detect the idea, we consider three strata, considering the non-depressed subjects, those who were moderately depressed and those who exhibited serious symp-



Figure 2. Relationship between weight loss, suicide attempt and depressive syndrome.

toms of this disorder. Table 1 can then be divided into three parts as shown in table 3.

As this hypothetical example can show, the association has «vanished» within each one of these strata, so that, once the depression is «controlled», it becomes clear that the presumed causal effect of the weight loss was false. In this illustration, the control of a single variable lead to ruling out the apparent causal relationship, but the contrary could occur: that, after the control of a confounding variable, «an association that was not seen in the beginning would «emerge.» On the other hand, what is typical is that several variables need to be controlled and the ideal is that this control be done simultaneously\*\*. The stratifying resource means loss of information when the variable to be controlled is quantitative (for example, age) and, on the other hand, it almost surely will be inapplicable when several variables must be controlled at the same time, since this would require the formation of perhaps dozens of strata and would thus require a sample of thousands of subjects.

In the decade of the 60's, the development of a very complex statistic technique was begun: logistic regression. Its noticeable synthesis power, versatility and modeling capacity, once rapid computers and advanced programs were available (essential in this case), made this statistical technique the most used in all contemporary biomedical investigation<sup>23</sup>.

Among its virtues is the fact that it totally substitutes the stratification approach due to Mantel and Haenszel in the sense that all that this approach can potentially supply remains as a specific case of that offered by the logistic regression. Furthermore, it exceeds the classical approach, since it makes it possible for the above described control process to be verified with several variables simultaneously with a relatively reduced sample and, in addition, it does so independently of the character of the variables to be controlled (these can be dichotomic, nominal, polytomic or continuous). For example, in a study where, in my opinion, the case-control methodology is used in an exemplary way<sup>24</sup>, due to the scrupulous management of the temporality and the careful selection of the 327 cases and 897 controls, depression is studied as a possible risk factor of ischemic heart disease. After find-

<sup>\*</sup> Exactly, it deals with a weighted average, for which a confidence interval can be calculated. We do not excessively complicate this article with formulas that the author can find in the statistical literature, or those that, simply, can be omitted (if there is an adequate computer program).

<sup>\*\*</sup> Unfortunately, there could be many other confounding factores, that are known or not (for example, genetic traits or death of one of the parents).

	Not depressed		Moderately depressed			Very depressed			
	WL	No WL	Total	WL	No WL	total	WL	No WL	Total
Cases	2	35	37	84	06	90	42	12	54
Controls	11	200	211	198	14	212	93	27	120
Total	13	235	248	282	20	302	135	39	174
	$OR_1 = \frac{ab}{bc} = \frac{(2)(200)}{(35)(11)} = 1,0$			$OR_2 = \frac{ab}{bc} = \frac{(84)(14)}{(6)(198)} = 1,0$		$OR_3 = \frac{ab}{bc} = \frac{(42)(27)}{(12)(93)} = 1,0$			

TABLE 3. Distribution of case-controls according to degree of depression and according to whether there was a considerable weight loss of not before the suicide attempt

WL: weight loss.

ing that there is a noticeable association between this factor and the disease, the logistic regression is used to corroborate that this association is maintained after simultaneously controlling the smoking habit, diabetes, hypertension and characteristics of the area in which the subject lived (socially depressed or not).

#### 3d. Biases

As is generically commented on in section 2d, there are many biases that may occur in any type of study. In this Section, we deal with two expressions that are typical of case-control studies: memory bias and selection bias.

To illustrate the first one, we imagine that the cases are individuals with Alzheimer's diagnosis and that the fact of having been treated with general anesthesia could be a risk factor is assessed<sup>25</sup>. We suppose that for the cases, the information is obtained from a relative and for the controls, from the individual himself. It is likely that when it is a patient, the family makes a special effort to remember any «exposition» that could have potentially affected him, especially an operation that has meant general anesthesia. On the other hand, a control may not have the same motivation to remember and it is likely that the information reported is less rigorous. In such case, it is said that the response is differential between the groups, which could introduce the previously mentioned memory bias. Probably, in this example, the association between exposition and outcome would be overestimated.

The most insidious bias that may occur in the casecontrol studies, however, is that of selection. To avoid it from appearing is, simply, the core methodological key to the case-control studies. The procedures to choose the controls may be adequately established as long as the criteria used to select the cases are clearly established. The theoretically most appropriate control group corresponds to the subpopulation of individuals who, having been at risk of suffering the study outcome, in the hypothetic case of having reached it, could have been chosen as cases.

The idea would be that any trait would have, at first, the same possibility for cases and for controls of being present in the beginning of the condition that the study attempts to reconstruct. That is, that no factor is more represented in one of the groups as a consequence of the way in which the groups were shaped. Stated in another way: a factor may appear more frequently among cases than among controls (which would make us think that it has a causal role), but it must be ruled out that this difference has been induced according to how the groups were selected.

This means that the controls should be selected from the same imaginary cohort that the cases come from. When the cases come from a well defined population in time and space, the selection of controls could be performed by a simple random sampling of this population. Thus, any bias due to this concept would be ruled out. However, this can only be rarely achieved.

In practical terms, it is crucial for the controls to be selected independently of whether they had been exposed or not exposed. If the exposition condition influences the possibility that an individual is included or not as a control in any way, a selection bias will be produced.

For example, we suppose that two factors that have been considered as a risk for Alzheimer's disease are assessed: consumption of raw meat and hypothryoidism. We suppose that all the subjects recorded as carriers of Alzheimer's disease in the hospital of a region in which this is the only service that sees such disease are taken as cases and that a sample of disease free patients that arrive to the emergency service of this hospital are taken as controls. In this situation, we can be generating a selection bias. In fact, consumption of raw meat is more frequent in rural areas than in the urban setting of the hospital. And it occurs that, while the service that sees the psychiatric problem receives patients from all the region, the emergency service only sees urban cases (the rural area emergencies go to centers in their own setting). Thus, the controls sample would have an underrepresentation of raw meat consumers beforehand in regards to the cases and this additional amount of raw meat consumers, that has nothing to do with the mental disease, could contribute to our erroneously giving causal weight to it (certainly, the causal character of the relationship between raw meat and the disease has been ruled  $out^{26}$ ). On the other hand, if hyperthyroidism occurs equally in the zone seen by the emergency service and all the region, the analysis in regards to this factor would

not be biased by the selection concept (the risk factor of hyperthyroidism condition is, in fact, consistent with recently found results<sup>27</sup>).

# 4. WARNINGS AND FINAL COMMENTS

To conclude this study, we will examine several insufficiently understood events and that are a not infrequent source of error.

# 4a. Prediction and causality

It is not difficult to find published studies in which, after finding a high OR, the condition of predictive is attributed to a variable. If a certain factor has a causal weight in an outcome, then it always has predictive value; and a factor that is not causal may even have such value. In general, there is no incorrection in it, as long as temporality has been contemplated (not necessarily plausibility or the effect of third factors). For example, weight loss may have predictive value for the effect of suicide although it does not have, as we have already seen, any «responsibility» in the fact. And in this quality, it can be of interest for the effects of prevention. However, what is especially interesting to stress is that the reciprocal is false. «Risk factor» and «predictive factor» are not synonymous: the fact that a risk factor is present or not may sometimes be useful for prediction, but a variable may make an important contribution to the effects of predicting, although by itself it is not a risk factor or a real component of the causal network. A factor can serve for prediction without making it, in any way, a causal agent.

Another closely related event with this subject deserves some consideration. It should be clear that incidence rates cannot be estimated with the case-control study data. For example, it would simply be nonsensical to conclude, from table 1, that 28 % (128/450) of those who have lost weight attempt suicide or that this occurs with 18% (53/294) of the adolescents who have not lost weight. Such impossibility is a direct consequence of the fact that, typically, the sample made up by cases and controls selected for the study is not representative of the population. In effect, if the outcome that is studied occurs in the population with a very low prevalence (for example, 2 per 1000), the case-control approach is the most adequate\*. In fact, the most attractive trait of the case-control studies consists in the fact that they are performed with very reduced samples. Thus, in this example, it would be typical that 200 cases and 400 controls, for example were worked with. On the other hand, while there is 1 subject with the disease in the population for every 499 who do not suffer it, there is only one patient for every 2 healthy subjects in the sample.

If a case-control study has been made and a variable or risk condition has been identified by the behavior, it could be asked: could the incidence rates be estimated using this information? That is, is it possible to estimate not only an approximate value of relative risk (the OR) but also the likelihood that a subject who has a risk condition would develop the disease? The answer is yes<sup>28</sup>, as long as there is a reliable estimation of the disease (or outcome) prevalence in the population. This is almost universally unknown, a reason why the interested reader is referred to an Appendix in which a problem solution is explained.

It can be stated that, for the same reason of lack of representativeness, the logistic function coming from a case-control study does not make it possible, as such, to directly estimate the likelihood that a certain outcome will occur for a subject who has a certain specific profile with predictive potentiality (regardless of whether the variables making up the model are causal or not). However, it is possible to use the logistic function, for this objective, if a correction of the estimation of one of the parameters is made, supposing again that we have the additional information for it<sup>29</sup>.

## Sample size

It is obvious that no analysis, regardless of the method chosen to carry it out, can be persuasive if it does not have sufficient data. Much has been written on how to reach a minimum sample size, in general, and even in relationship with the psychiatry studies<sup>30</sup>. In fact, much has been written, but not many different things; the articles or book chapters often seem to be cloned from each other. There are also many computer programs designed to solve this problem. EPIINFO 6,0 and EPIDAT 3,0 are, among them, especially flexible and friendly. Thus, we will not spend much time to give formal guidelines to confront this important problem. I consider it appropriate, however, to make a very general consideration.

The problem of the sample size has been historically managed idyllically: the official discourse of statistics eliminates the discussion on the noticeable obstacles that comes from applying the formulas and substitutes it with operative guidelines of doubtful applicability. Some authors are emphatic on stating that there is a false panacea behind the formulas. The famous epidemiologist Kenneth Rothman; for example, writes<sup>31</sup>:

In short, the problem of determining the most adequate sample size does not have a technical nature, susceptible to being solved by calculations but rather it must be confronted by judgment, experience and intuition.

However, in general, this reality is overlooked or buried under standard attitude that almost all repeat meaninglessly. It is useful to understand that whatever the size of the sample, both the confidence intervals as well as

<sup>\*</sup>A prospective study, for example, would require a sample of many thousands of subjects to be able to count on minimally exact incidence rate estimations.

# APPENDIX. Estimation of incidence rates from a case-control study

If we call  $p_1$  and  $p_2$  for the rates exposed between the cases and the controls, respectively, and f the general prevalence of the outcome in the population, then it can be demonstrated by the bayes theorem that the development rates of the diease (or appearance of the outcome) between those exposed and not exposed respectively come from the following formulas:

$$t_1 = \frac{p_1 f}{p_1 f + p_2 (1 - f)} \qquad t_2 = \frac{(1 - p_1) f}{(1 - p_1) f + (1 - p_2) (1 - f)}$$

For example, we consider the case of depression and suicide attempts summarized in table 2. If it is admitted that the suicide attempt rate between adolescents of the community is 2 per 1,000 (f = 0.002), remembering that

$$p_1 = \frac{144}{181}$$
 and  $p_2 = \frac{211}{543}$ , we would have:

 $t_1 = 0.00408$  and that  $t_2 = 0.00067$ . Remember that, in this example, the OR is equal to 6.12 virtually equal to the rate ratio (RR) is

$$\frac{t_1}{t_2} = \frac{0.00408}{0.00067} = 6.09$$
, as was to be expected

The same could be done using table 1 for weight loss. We remember that it makes sense to calculate these rates independently of whether it deals with a true risk factor or not; in short, for the effects of the secondary prevention, it could be useful, for example, to have an estimation of the likelihood that an adolescent who has lost weight will try to commit suicide.

the *p* values can be calculated *a posteriori*. Consequently, the most general suggestion that can be given on this subject is that, without scorning the orientative value that the results arising from the application of formulas may have, the sample size is chosen, above all, using common sense, observing what is done in the literature and taking the resources available into account. In chapter 11 of a book published in 1997, I give my points of views on this polemic matter in detail<sup>32</sup>.

A related problem concerns the relationship that must be established between the number of cases and controls. The most natural is that they are equal. Although the controls are generally more available than the cases, identification and questioning of the controls are often a long and burdensome process, a fact that can advise equality. However, such equity is far from being necessary, and it may even be convenient to have more cases than controls. In fact, when the number of cases is limited, a natural alternative is that of compensating this circumstance in some way, increasing the number of controls. The increase in the number of controls per case is useful until a 4 to 1 ratio is reached. From here on, not much is gained by increasing the sample size at the cost of increasing the control number<sup>33,34</sup>.

#### 4c. Stepwise regression

Any user of statistical programs knows the existence of a resource aimed at subselecting a multiple regression model (that is, an algorithm that makes it possible to reject some variables initially considered to construct a «final model»). Unfortunately, the comfortable resource of stepwise regression not only is sterile for the causality analysis but also can be simply harmful, so that it should be directly avoided. It only makes sense when we want to construct a predictive model from a case-control study (which would require making the correction mentioned in section 4a).

Sometimes the procedure is used to discover which are the causal variables and rule out by behavior those that are not. Other times it is used, but it is not clear why. In a case-control study to identify risk factors to become an antidepressant consumer<sup>35</sup>, a typical maneuver is done. With the variables that exhibit a bivariate level relationship with the fact of being a case or control, a logistic regression was adjusted and then a stepwise method was applied. According to the words of the investigators: after applying this procedure «the fact of having lived some relevant personal event and presenting values on the Zung\*\* scale that are superior to 50» became part of the equation. A «final» model has been constructed; it is now appropriate to ask the question: and what does that mean? It would give the impression that the application of this recourse has been made an objective by itself. At least in this specific study, no attempt is made to draw conclusions (the fact remains detached and disconnected in the discourse), since any response on causality would lack meaning. In effect, it would be impossible to rule out that, if the Zung scale «remains» in the model, there is a certain variable associated with this scale that would be statistically redundant (in virtue of which, therefore, it would be outside of the function obtained with the stepwise) but could have an important causal link with the consumption of psychodrugs (especially in this case, in which that «demonstrated» through the regression is no less than that antidepressant consumption is associated with being depressed). It is not difficult to find examples in the present literature in which, erroneously, conclusions are obtained<sup>36</sup>.

The fundamental problem arises from the presumptuous and also naive interpretation that is generally obtained from the result arising from the stepwise regression. Its use with explanatory objectives (identify causal factors) is absurd, since the algorithmic selection of models cannot prevent the results from being obtained from mere statistical concomitances (in fact, they are based on this), or distinguishing between the causal type associations and those due to third factors involved in the condition. Consequently, although the logistic re-

<sup>\*</sup>The Zung scale is an indicator proposed in 1990 to measure depression, which is obtained after a self-applied questionnaire.

gression model may be of extraordinary interest to help understand the biological and social conditions after a case-control study, as was stated in section 3c, the subselection algorithmic procedures of variables to make a «final» explanatory model are totally inadmissible.

## 4d. Confidence intervals versus hypothesis tests

In the first half of the XX century, the health care investigators generally did not have a good grasp on the methods that made it possible to quantify the evidence and thus complement the verbal reasonings of their investigators. In this time period, however, the so-called hypothesis tests slowly developed and became consolidated. In fact, in the 1920's, an approach that was invented by Ronald Fisher was born. It measures the degree of incompatibility of the data with a hypothesis and, with it, the famous p values. Some years later, Jerzy Neyman and Egon Pearson proposed a procedure focused on the choice between two hypotheses. Shortly after, the hybrid resource arising from the fusion of both contributions began to be anonymously managed<sup>37</sup>. The hypothesis tests, fedback by the growing universalization of the access to powerful computer resources and the development of many statistical software programs (SPSS, SAS, BMDP, EPINFO, etc.) that offer these hypothesis tests as one of their main attractions, consolidate their almost universal presence in biomedical investigation.

However, parallelly, many and persuasive objections to the use of the hypothesis tests were accumulating. The problems mentioned have various natures and we will not go into them now. However there is no doubt that they are extremely important and have been established over more than 40 years, both in many scientific articles<sup>38-42</sup>, as well as in textbooks<sup>43-46</sup>. To see a very complete summary with almost 400 references that include the last 40 years, especially the last 10, see the impressive study of Raymond Nickerson<sup>47</sup>.

As a consequence, since several years ago, several important journals in the international scientific production growingly tend to reject papers in which only tests of this type appear. For example, the British Heart Journal announced in a 1988 editorial that they would adhere to the requirement which had appeared in the British Medical Journal since 1986 when Gardner and Altman exhorted the authors to use confidence intervals instead of significance tests. This posture has been shared by such important journals as Lancet, Annals of Internal Medicine and American Journal of Public Health<sup>48</sup> and, finally, was adopted by the Vancouver Group<sup>49</sup> (International Committee of Medical Journal Editors), in whose technical requirement section dedicated to the use of statistics, it is recorded textually: «Avoid relying solely on statistical hypothesis tests, such as the use of p values, which fail to convey important information about the effect size». Other regulations type bodies of scientific activities have been slowly declaring themselves in the same direction (See, for example, the Wilkinson's study<sup>50</sup> on the recommendations of the Force on Statistical Inference created by the Association of Psychologists of the United States on the use of statistics in psychological investigation).

The intrinsic weaknesses of the method, together with the ease with which it can be erroneously interpreted, answering the call that the common use that it is generally given daily and the support that the serious objections have received from the authorities and committees as those mentioned, lead to predict that sooner than later, a new era will be consolidated in which the use of the confidence intervals and Bayesian statistics will displace the hypothesis tests.

Although, according to such reality, there are increasingly more authors and medical journals that decide to avoid the use of this resource, it is true that the hypothesis test continues to be of daily usage<sup>51</sup>, even in journals formally attached to the Vancouver Group. It is obviously reasonable to suggest to the investigators that they remain attentive to the evolution of this process and, in fact, that they abide by the guidelines explicitly adopted by the journals in which they attempt to publish their results.

# 4e. Language and conclusions of a case-control study

After finding an OR that is noticeably larger than the unit, many investigators «conclude» that the variables to which this OR corresponds (or those that remain in the model if adjusted to a logistic regression, that is stepwise or not stepwise) are associated with the phenomenon that is studied (typically a disease), thus avoiding the commitment to give an opinion on whether such variables are the cause or not of the phenomenon in question. In the same way that it makes no sense to establish the association as a purpose (see section 2a), it makes no sense to convert the statement that the two variables are associated into a conclusion, because the latter should be the result of an intellectual process that is qualitatively superior to the mere phenomenological quantification that the former represents and, finally, because the conclusions should be answers to the questions that are formulated, and among these, it is not legitimate to ask if there is an association or not.

On the other hand, the terminological confusion existing is noticeable. In an article on the contemporary use of the stepwise logistic regression<sup>36</sup>, it was reported that when papers published in the Spanish journal of *Medicina Clinica* are reviewed, we find the most diverse expressions to qualify or name the variables that are being studied within the same text and preceding the expression that refer to the disease or problem studied. Besides the fact that these expressions are occasionally intrinsically questionable, the fundamental problem is that they are used in an apparently irreflexive way (in one individual article, they are even sometimes called in one way and other times in another, in spite of the different semantic meanings that they have). Sometimes reference is made implicitly or explicitly to causal relationships («factors that have an effect on», «risk factors», «variables influencing in», «variables explaining», «more decisive factors in»). There are no objections to these expression if, in fact, they are in studies in which an attempt is made to identify explanatory factors of the condition studied. On other occasions, reference is made to their predictive potentiality («predictors», «factors prognostic of», «predictive variables»), which also may be correct in some contexts. Other expressions are directly incorrect; such is the case of the «factors predicting risk» (since it is not the risk that is predicted but rather an outcome) or of «mediators» (a totally undefined term and therefore, ambiguous). Well-known references to the association often appear: «variables associated with», «variables related with». Finally, expressions, such as «variables statistically responsible for», may be found, which, in my opinion, are directly preposterous.

# 5. FINAL CONSIDERATION

As a last note, I feel that it is adequate to stress that the performance of a methodologically solid case-control study entails a true challenge that cannot be solved following a recipe. The practical advantages that this approach has in regards to experimentation and prospective observational studies are significant, but require a high dose of critical reflection and always mean an exercise of creativity. For many, this may add an additional attraction, especially in a field such as psychiatry, in full development and with so many problems open.

## **REFERENCES**

- 1. Hill AB. Environment and disease: association or causation? Proceedings of the Royal Society of Medicine 1965; 58:295-300.
- Bofill J. Integración de los criterios de Sir Austin Bradford Hill a los criterios utilizados tradicionalmente para el estudio de la casualidad en valoración del daño corporal. 1999 (disponible en http://www.la-plaza.com/vdc/BD/ doc/revisiones/index.html).
- Silva LC. La homeopatía, una crónica bicentenaria. Revista Habanera de Medicina 2002;1(2). Disponible en http://www.sld.cu/instituciones/iscmh/rhabcm/rhbcm1/ Articulos/articulo\_luisc.htm.
- 4. Davidson J, Gaylord S. Meeting of minds in psychiatry and homeopathy: an example in social phobia. Alternatives Therapies in Health and Medicine 1995;1(3):36-43.
- 5. Carlston M. The mechanism of homeopathy. All that matters is that it works. Alternat Ther Health Med 1995;1: 95-6.
- Silva LC, Benavides A. Causalidad e inobservancia de la premisa de precedencia temporal en la investigación biomédica. Rev Methodol (Belg) 1999;7:1-11.
- Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet 2002;359:431-4.
- Schlesselmann JJ. Case-Control studies. New York: Oxford University Press, 1982.

- Breslow N, Day NE. Statistical methods in cancer research Vol 1: The analysis of case-control studies 1980; IARC Scientific Publication No. 32 (Lyon: IARC).
- 10. Szklo MF, Nieto J. Epidemiology: beyond the basics. Gaithersburg: Aspen, 2000.
- 11. Cornfield J. A method of estimating comparative rates from clinical data. Application to cancer of the lung, breast and cervix. J Nat Cancer Instit 1951; 11:1269-75.
- 12. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. BMJ 1952;2:1271-86.
- 13. Silva LC. Diseño razonado de cuestionarios y muestras para la investigación sanitaria. Madrid: Díaz de Santos, 2000.
- 14. Zornberg GL, Jick H. Antipsychotic drug use and risk of first-time idiopathic venous thromboembolism: a case-control study. Lancet 2000;356:1219-23.
- Gardner MJ, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. BMJ 1986;292:746-50.
- Fairburn CG, et al. Risk factors for anorexia nervosa: three integrated case-control. Comp Arch Gen Psychiatry 1999;56:468-76.
- 17. McCreadie RG. Use of drugs, alcohol and tobacco by people with schizophrenia: case-control study. Br J Psychiatry 2002;181:321-5.
- Vilalta J, Llinàs J, López S. Depresión y demencia: estudio caso-control. Rev Neurol 1999;29(7):599-603.
- 19. Kresnow M. An unmatched case-control study of nearly lethal suicide attempts in Houston, Texas: research methods and measurements. Suicide and Life-Threatening Behavior 2001;32(Suppl):7-20.
- MacDonald S, et al. Nithsdale schizophrenia surveys 24: sexual dysfunction: case-control study. Br J Psychiatry 2003; 182:50-6.
- 21. De Irala J, Martínez MA, Guillén F ¿Qué es una variable de confusión? Med Clin (Barc) 2001;117:377-85.
- 22. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Nat Cancer Institute 1959;22(4):719-48.
- 23. Silva LC, Pérez C, Cuellar I. Uso de métodos estadísticos en dos revistas médicas con alto factor de impacto. Gaceta Sanitaria 1995;9(48):189-95.
- 24. Hiipsley J, Fielding K, Pringle M. Depression as a risk factor for ischaemic heart disease in men: population based case-control study. BMJ 1998; 316:1714-9.
- 25. Broe GA, Henderson AS, McCusker E, et al. A case-control study of Alzheimer's disease in Australia. Neurology 1990; 40:1698-707.
- 26. Chandra V, et al. Case-control study of late onset «probable Alzheimer's disease». Neurology 1987;37:1295-300.
- 27. Breteler MMB, et al. Medical history and the risk of Alzheimer's disease: a collaborative re-analysis of case-control studies. Internat J Epidemiol 1991;20(Suppl 2):836-42.
- 28. King G, Zeng L. Estimating Risk and rate levels, ratios, and differences in case-control studies. Statistics in Medicine 2002; 21:1409-27.
- 29. Silva LC. Excursión a la regresión logística. Madrid: Díaz de Santos, 1995.
- Gómez C, Sánchez R. Cálculo del tamaño de la muestra en psiquiatría y salud mental (principios básicos para su estimación). Rev Colomb Psiquiatr 1998;27(2):131-42.
- 31. Rothman JK. Modern epidemiology. Boston: Little, 1986.
- 32. Silva LC. Cultura estadística e investigaciones en el campo de la salud. Madrid: Díaz de Santos, 1997.
- Lasky T, Stolley PD. Selection of cases and controls. Epidemiol Rev 1994;16(1):6-17.
- 34. Gordis L. Epidemiology. Philadelphia: WB Saunders, 1996.

- 35. Rispau A. Factores de riesgo asociados al consumo de antidepresivos. Atenc Prim 1998;22(7):78-83.
- Silva LC, Barroso I. Selección algorítmica de modelos en las aplicaciones biomédicas de la regresión múltiple. Medic Clín 2001;116:741-5.
- 37. Goodman SN. Toward evidence-based medical statistics (I): The p value fallacy. An Intern Med 1999;130:995-1004.
- Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. Psychologic Rev 1963; 70:193.
- Carver RP. The case against statistical significance testing. Harvard Educ Rev 1978;48:378-99.
- Berger JO, Berry DA. Statistical analysis and the illusion of objectivity. Am Sci 1988;76:159-65.
- Tamhane AC. Reseña sobre el libro Bechhofer RE, Santner TJ, Goldsman DM. Design and analysis of experiments for statistical selection, screening and multiple comparisons, John Wiley (New York), 1995. Technometrics 1996;38: 289-90.
- Johnson DH. Hypothesis testing: statistics as pseudoscience. Fifth Annual Conference of the Wildlife Society, Buffalo, New York, 26 September, 1998.

- 43. Jeffreys H. Theory of probability, 3rd ed. Oxford: University Press, 1961.
- 44. Lindley DV. Introduction to probability & statistics. Part 2: Inference. Cambridge: University Press, 1970.
- 45. O'Hagan A. Kendall's advanced theory of statistics. Vol 2B: Bayesian Inference. London: Arnold, 1994.
- 46. Lee PM. Bayesian statistics: an introduction, 2nd ed. London: Arnold, 1997.
- 47. Nickerson RS. Null hypothesis significance testing: a review of an old and continuing controversy. Psychological Methods 2000;5(2):241-301.
- Evans SJW, Mills P, Dawson J. The end of the p value? Br Heart J 1988;60:177-80.
- Comité Internacional de Directores de Revistas Médicas. Requisitos uniformes de los manuscritos enviados a revistas biomédicas (mayo 2001).
- Wilkinson L. Task force on statistical inference. Statistical methods in psychology journals. Am Psychologist 1999; 54:594-604.
- 51. Sarria M, Silva LC. Empleo de las pruebas de hipótesis en la literatura biomédica iberoamericana actual. Revista Cubana de Salud Pública, 2003 (in press).