

How to Evaluate and Interpret a Scientific Article About Therapy Or Therapeutic Procedures?

Cómo Evaluar e Interpretar un Artículo Científico sobre Tratamiento o Procedimientos Terapéuticos

Tamara Otzen^{1,2}; Carlos Manterola^{1,3}; Maricela García-Domínguez⁴ & MInCir Group

OTZEN, T.; MANTEROLA, C.; GARCÍA-DOMÍNGUEZ, M. & MINCIR GROUP. How to evaluate and interpret a scientific article about therapy or therapeutic procedures?. *Int. J. Morphol.*, 35(2):776-784, 2017.

SUMMARY: Regardless of the subject area and design used, it has been verified that between 40% and 60% of the studies published in biomedical journals are articles about therapy or therapeutic procedures (TP). Anyone writing a manuscript related to therapy or TP or reading an article of this type must demand at the very least a clear, precise and concise objective with respect to the research conducted, explicit mention of the design used with the respective inherent methodological details, and the mention and execution of statistical tools and the measures of association, or at least the numbers needed to calculate these values. The aim of this manuscript is to present a synthesis of the fundamental elements for the correct writing, reading and assessment of such articles, regardless of the disciplinary area in which the research originated.

KEY WORDS: "Therapeutics"[Mesh]; "therapy "[Subheading]; "Risk"[Mesh]; "Clinical Trial "[Publication Type]; "Cohort Studies"[Mesh]; "Meta-Analysis "[Publication Type].

INTRODUCTION

Clinical research articles can be grouped according to the type of scenario addressed or of the research question to answer. Thus, we have articles about therapy, prevention, harm and etiology, prognosis, diagnosis, prevalence and differential diagnosis as well as economic analysis articles (Manterola, 2009; Manterola *et al.*, 2014).

It seemed relevant to us to write this paper considering the high frequency of publications related to therapy or TP because, regardless of the subject area and design used, it has been verified that between 40 % and 60 % of the studies published in biomedical journals are articles about this type of scenario (Manterola *et al.*, 2006a; Manterola *et al.*, 2006b; Manterola & Grande, 2010), a more than good enough reason to substantiate their correct reading and assessment.

Ideally, reporting results from studies on therapy or TP should arise from valid and reliable studies with a good level of evidence and a degree of recommendation; i.e., from systematic reviews (SR) of individual with homogeneity

randomized clinical trials (RCT), and controlled, masked and with narrow confidence interval RCT; designs that represent evidence levels 1a and 1b respectively and degree of recommendation A (Manterola *et al.*, 2006a; Manterola *et al.*, 2006b; Manterola & Grande, 2010; Manterola *et al.*, 2014). However, the reality is quite different, and the publications on therapy and TP include a wide variety of forms and depth: forms due to the diversity of existing designs that range from the classic observational to the experimental, and depth because in spite of finding a greater or lesser approach to a design in most publications, it is also frequent to find weaknesses that threaten the validity and reliability of their results. Thus it has been determined that around 80 % of articles referring to therapy or TP are studies with evidence level 4 (reports and case-series and poor quality cohort studies or retrospective cohorts) of low methodological quality that contain serious methodological shortcomings (Manterola *et al.*, 2006a; Manterola *et al.*, 2006b; Manterola & Grande, 2010). These data reinforce even more the idea of assessing scientific articles

1 Center of Morphological and Surgical Studies (CEMyQ), Universidad de La Frontera, Temuco, Chile.

2 Facultad de Ciencias de la Salud, Universidad de Tarapacá, Arica, Chile.

3 Department of Surgery, Universidad de La Frontera, Temuco, Chile.

4 Universidad Nacional Autónoma de México - Unidad León, México.

appropriately through a critical reading for which specific tools are available.

The key points of a manuscript related to therapy or TP are the title, the research question, the aim of the study, the design used and the respective level of evidence, the statistical tools and the measures of association used.

The aim of this article is to provide basic methodological concepts that must be considered when a study on therapy or TP is assessed and interpreted.

THE TITLE

It is essential to have a suggestive title that piques the reader's curiosity so as to motivate them to read the abstract and then the text. Its main function is to accurately describe the content of the manuscript. With the fewest words possible it must be able to outline the existing uncertainty with respect to the subject matter while simultaneously explaining the type of study (Manterola *et al.*, 2007).

Sometimes the authors select a title that contains these features. This occurs in particular when the writing is guided by special standards such as CONSORT, QUOROM, STROBE, TREND, etc. (Moher *et al.*, 1999; Moher *et al.*, 2001; Des Jarlais *et al.*, 2004; Vandembroucke *et al.*, 2007; Zwarenstein *et al.*, 2008); for example, "Surgery for morbid obesity: selection of operation based on evidence from literature review" (Manterola *et al.*, 2005a) or "Open versus laparoscopic resection in non-complicated colon cancer. A systematic review" (Manterola *et al.*, 2005b). In both cases, the research question and the aim of the study are set forth more or less implicitly as well as the design used (a SR). "One- and ten-year outcome of laparoscopic anterior 120° versus total fundoplication: a double-blind, randomized multicenter study" (Djerf *et al.*, 2016): this example expresses the aim of the study as well as the design used (a RCT).

On other occasions, the authors opt for a title that contains these characteristics only partially. "Efficacy of Nissen fundoplication versus medical therapy in the regression of low-grade dysplasia in patients with Barrett esophagus: a prospective study" (Rossi *et al.*, 2006) or "Intra-oesophageal acid suppression in complicated gastro-oesophageal reflux disease: esomeprazole versus lansoprazole" (Frazzoni *et al.*, 2006): in both examples, the question and the aim of the study are expressed more or less precisely, but not however the design used, which could be a SR, a RCT, a prospective or even a retrospective cohort study.

But the most common is when a simple title is chosen which does not clarify what the authors are trying to report. For example, "Multivisceral echinococcosis: concept, diagnosis, management" (Grozavu *et al.*, 2014) or "Hepatic Hydatid Cysts

Causing Biliary Obstruction" (Caballero-Mateos *et al.*, 2017): in both cases the title is a mere description of a phenomenon so there is no way to suppose what type of design was used. In the first example at least the word management is mentioned, without specifying to what it refers; but in the second it is impossible to even suppose that this is an article about therapy or TP.

RESEARCH QUESTION AND OBJECTIVES

When it is time to decide whether the study is about therapy or TP, the research question, when the author provides it, gives the suitable information. Not being indicated (the most common), the aim of the study can help to understand the nature or clinical scenario of the article.

With respect to the clinical question, it must be considered that a structured approach to its concept is the first step to designing a study. Above all it must be precise and focused on the issue raised. The best way to do this is by ensuring a series of basic components are incorporated from the scheme known as PICO: The problem that creates uncertainty in a Patient/Population; the Intervention to be assessed; the Comparator for the study intervention that must be considered according to the question and the design used, and the Outcome that will be measured from the intervention applied (Fisterra.com, 2017). For example, in the case that the effectiveness of the gastrectomy and D1 regional lymphadenectomy with adjuvant chemoradiotherapy for the therapy of resectable gastric cancer is to be assessed, we will have to describe the study population, the intervention, the comparator or alternative therapy, and the period of time if necessary in sufficient detail (Table I).

In this situation, the question could be written as: What is the best therapy for resectable gastric cancer in terms of 5-year survival between a gastrectomy and D2 regional lymphadenectomy and gastrectomy and D1 regional lymphadenectomy with adjuvant chemoradiotherapy?

The lack of clarity and precision of a question is among others things associated with a high probability of error in calculating the sample size needed for the study and therefore also with the certainty of the sample estimation, the precision of the inference, the statistical power or the ability to detect differences if they exist, etc.

On the other hand, the objective is the axis around which the structure of the study is constructed. If this is not clear, precise and concise, it will be difficult to

Table I. Examples of the PICO strategy for the generation of research questions.

P (patient)	I (intervention)	C (comparator)	O (outcome)
Patients with esophagitis due to repeated/chronic gastroesophageal reflux	Laparoscopic Nissen fundoplication	Permanent administration of Lansoprazole	of ✓ Control of symptoms ✓ Control of the appearance of ✓ erosive lesions of the esophageal mucosa
What is the best therapy for the control of symptoms of gastroesophageal reflux disease and the development of secondary esophagitis between laparoscopic Nissen and Lansoprazole?	single, Administration of Albendazole	Laparoscopic pericystectomy	✓ R recurrence ✓ M morbidity ✓ C costs involved
What is the best therapy for single, uncomplicated hepatic hydatidosis in terms of recurrence between the administration of Albendazole and laparoscopic pericystectomy?	Gastroectomy and lymphadenectomy with chemoradiotherapy	D ₁ regional Gastrectomy and D ₂ regional lymphadenectomy	✓ 5 - year survival ✓ M morbidity ✓ M mortality ✓ A dverse effects ✓ R recurrence ✓ C osts involved
Patients with resectable gastric cancer	Gastroectomy and lymphadenectomy with chemoradiotherapy	D ₁ regional Gastrectomy and D ₂ regional lymphadenectomy	✓ 5 - year survival ✓ M morbidity ✓ M mortality ✓ A dverse effects ✓ R recurrence ✓ C osts involved
What is the best therapy for resectable gastric cancer in terms of 5-year survival between a chemoradiotherapy and gastrectomy and D ₁ regional lymphadenectomy with adjuvant chemoradiotherapy and gastrectomy and D ₂ regional lymphadenectomy?	Gastroectomy and lymphadenectomy with chemoradiotherapy	D ₁ regional Gastrectomy and D ₂ regional lymphadenectomy	✓ 5 - year survival ✓ M morbidity ✓ M mortality ✓ A dverse effects ✓ R recurrence ✓ C osts involved

discern the type of study; furthermore, in such a situation (unfortunately very frequent), the writing of the objective will only add greater uncertainty and doubts with respect to the selection of the study population, the sample size needed, the study variables and the subsequent statistical analysis. A frequent problem in biomedical articles is that the research aims are usually vague and inexact, or sometimes they do not even feature in the manuscript (Manterola *et al.*, 2006a; Manterola *et al.*, 2006b). Thus, imprecise aims such as, “To evaluate the short- and long-term outcomes of liver resection for caudate lobe hepatocellular carcinoma” (Liu *et al.*, 2010) pose the disadvantage of not making it clear what results are going to be evaluated (overall or disease-free survival? recurrence? morbidity? etc.); nor is it about patients with hepatocellular carcinoma in general or some subtype or advanced stage of the disease, or even about the type of resection that was performed (partial, total or extended). One option to improve this situation could be: “to evaluate the results of the total lobectomy of the caudate lobe in patients with stage II and Child-Pugh A hepatocellular carcinoma in terms of overall survival and recurrence”. This is because, in this example, patients are routinely assessed with different types of histology, stages, hepatic functional reserve, type of resections, etc., and despite all this, conclusions are drawn that can apply to some scenarios but not all.

TYPES OF DESIGN AND THEIR LEVEL OF EVIDENCE

Considering the primary standpoint of the question on therapy or TP, how does the therapy change the clinical course of the disease? It may be supposed that responding to it involves a series of variables to consider, in addition to the time, i.e., the follow-up period, from when the study therapy is administered until a change occurs in the clinical course of the disease.

Any article must declare explicitly the design used in the study, and articles referring to therapy or TP are not an exception.

The study designs that involve a follow-up time are cohort studies (among the observational studies) and the RCT (among the experimental studies), both with all their respective variants (Manterola & Otzen, 2014; Manterola & Otzen, 2015; Manterola *et al.*, 2009). Nevertheless, it has been verified that between 70 % and 80 % of articles on therapy or TP are reports and prospective and retrospective case series (Manterola *et al.*, 2006a; Manterola *et al.*, 2006b; Manterola & Grande, 2010).

However, if we return to the question, how does the therapy change the clinical course of the disease? The following question is asked implicitly: is the study therapy the cause of the change in the clinical course of the disease? In other words, is there a causal association between the therapy and the disease? This is therefore about the Cause and Effect relationship. In this respect, a connection can be defined as the statistical dependency between two or more factors, where the occurrence of one factor increases (or decreases) as the other varies. But its presence does not mean that the relationship is necessarily cause-effect, then the primary aim when assessing a study on therapy or TP is to judge when a therapy -disease relationship is causal.

A causal association is one where the change in the frequency and quality of a therapy or TP results in a corresponding change in the frequency of the disease. This way, judging when the association is causal extends beyond the validity of the results of any study and includes the consideration of the epidemiological data as well as the biological credibility of the hypothesis.

If in a study on therapy or TP it is determined that chance, bias and confounding are unlikely to explain the change in the course of the disease, then it may be concluded that there is a valid statistical association. It is therefore necessary to consider whether the relationship is cause and effect, since the presence of a statistically valid association does not imply causality.

There are criteria that can help in the causality judgments, including the force of association, the biological credibility of the hypothesis, the consistency of the findings as well as other data related to the time sequence and the presence of a dose-effect relationship (Hennekens & Buring, 1987; Feinstein, 1995; Kelsey *et al.*, 1996). The basic reasoning to establish a causal relationship is the sequence of events, i.e., that the cause is present before the effect is produced. However, prior to establishing that two or more factors have a cause-effect relationship, it must be shown that the link between them is valid; this means that a valid association is a real or true association, where the effect of chance and bias is minimal.

Consideration must also be given to random error or chance because this is inherent to all observations and can be assessed by applying a test of statistical significance, the objective of which is the p value (Manterola & Pineda, 2008).

Another instrument to assess the influence of chance is the determination of the 95 % confidence interval (95 % CI). In statistical terms, this is the interval of numerical values in which the population value that is being estimated is found

with a 95 % confidence level (Riegelman, 2013). The 95 % CI of a measure of association contains the real value of this measurement with a 95 % certainty. It should be borne in mind, however, that the information given by the p value and 95 % CI is complementary. It is also important to remember that the p value and the 95 % CI are dependent upon the sample size such that the smaller the sample, the greater the p value and the wider the interval (Dawson & Trapp, 2005; Manterola & Pineda, 2008).

Another concept worth noting is observational and involuntary errors. These appear when the compared components are not sufficiently similar. Therefore, they can occur at any stage of the process of evaluating an association, emphasizing selection, measurement and confounding biases (Fletcher *et al.*, 2002; Manterola & Otzen, 2015).

In short, the effects of chance and bias on the evaluation of an association are related to the methodological quality of the study. The types of bias considered bring about a distorted comparison within the cause-effect reasoning model. Despite a good internal comparison, the results may not be generalizable or extrapolated to a different scenario; this occurs when the study groups have a distorted selection of the population they supposedly represent.

For all these reasons, it may be said that the level of evidence of clinical research designs is directly related to the force and size of the causal association on the understanding that these tell us about the proximity to the real value of the estimation. From this point of view, the best level of evidence for studies on therapy or TP are found in SR with or without a RCT meta-analysis (Manterola, 2009; Manterola *et al.*, 2014), followed by the individual RCT with narrow Confidence Interval and observational studies (cohort studies, case-control studies, case-series, etc.) (Manterola *et al.*, 2014).

So, it is essential that RCT be planned with random allocation and masking. The advantage of the random allocation process is that the variables related to prognosis, known and unknown, are distributed similarly among the study groups in such a way that any difference recorded can be attributed to the different therapy modalities received by one group or another. The advantage of the masking process is that it allows for a more objective measurement in such a way that neither the researcher nor the study subject know what the intervention is that has been assigned to each group, thereby avoiding a bias by either of these two. This is a situation that in many cases is not possible as a result of either ethical problems or feasibility. A typical example is when an attempt is made to compare results of laparoscopic surgery and conventional surgery. In these cases the

researcher cannot avoid informing the patient that one technique or another was applied.

On the other hand, in the case of analytical observational studies, such as cohort studies and case-control studies (Manterola & Otzen, 2014), since the researcher does not control the variables as in an experimental situation, it is not possible to be ensure that other factors do not differ between one study group and another. Therefore, the estimation of the effect of the therapy could be biased by prognostic factors, which is why the level of evidence of such studies is lower (Manterola *et al.*, 2014). However, with these studies it must be emphasized that the main difference between cohort studies and case-control studies does not lie in the time sequence of the investigation, but rather in the selection criteria of the study populations; nevertheless, and given that part of the existence of an “event of interest” or of a disease, there is difficulty in ensuring a logical time sequence where the exposure precedes the effect (Manterola & Otzen, 2014). Both designs have the disadvantage of being subject to several biases, including selection, interviewer and memory biases, among others.

In addition, we can find the evaluation of therapy or TP performed with descriptive observational studies; i.e., those designs in which there are no control groups or reference groups with which to compare (some cross-sectional studies, correlational studies, case series and case reports). These types of studies are, as previously mentioned, the most common in scientific journals (Manterola *et al.*, 2006a; Manterola *et al.*, 2006b). Cross-sectional studies provide a snapshot of the coexistence of exposure and effect, and have the same methodological limitations as the case-control study; in addition, they have greater difficulty demonstrating the time sequence of cause and effect (Hernández & Velasco, 2000). Finally, the case series and case report are useful for describing the results observed in a patient or a group of patients with a similar health problem, considering that they deal with experiences limited to the observations made by a researcher or a group of researchers deprived of a control group or comparison, a situation associated with a very high likelihood of every type of bias, this gives such studies a low level of evidence in all the classifications in use today (Manterola, 2014).

APPLICATION OF STATISTICAL TOOLS AND MEASURES OF ASSOCIATION

It does not seem necessary at this point to enter into detail about all the statistical tools available and which can be used in articles related to therapy or TP.

Nevertheless, it seems reasonable to remember that

there is a “central thread” that will always begin with the description of the study sample, i.e., applying the so-called descriptive statistics with a calculation of percentages, measures of central tendency (average, median, mode) and dispersion (range, variance and standard deviation), determination of prevalence, estimation of 95 % CI, survival curves, etc. Later, the bivariate analyses are applied, using the well-known Pearson’s Chi² and Fisher’s exact test for the comparison of the values of frequency between categorical variables, parametric tests like the t-test or student’s t and ANOVA among others for the comparison of averages, non-parametric tests for the comparison of variables of skewed distribution, and multivariate analyses using linear, logistic or ordinal regression models as appropriate.

All this will depend on the type of design, the characteristics of the population and the variables with which the research group is working.

Yet it seems fundamental to stress the notion that a p value that is statistically significant or not must be assessed in each context, because it is nothing more than a value that may be “statistically significant” or not and is not necessarily associated with the multi-factor dynamics of the clinic. It is not unusual to observe in some articles that a p value of 0.045 is considered “statistically significant”, which strictly speaking it can be. However, before ensuring it, the population characteristics, the sample size used for the study, which statistical tools were used, etc., need to be assessed. Subsequent to all the above, and if dealing with RCT, cohort studies and case-control studies, the magnitude of the effect of the study therapy or TP must be assessed in terms of the standard in use or another one, for which there are some tools to compare the risk to the group receiving the intervention vs. the risk to the control group. These are the so-called measures of effect (based on the quotient) and the measures of impact (based on the difference).

The measures of effect are the estimation of the “relative risk” (RR) and the “odds ratio” (OR). The use of one or the other will depend on the study design; thus, if it is a case-control study, OR should be applied; and if it is a RCT or a cohort study, RR should be applied (Manterola & Otzen, 2015).

The OR is the quotient between the likelihood that the event will occur and the probability that it will not (odds); therefore, it indicate how likely the event will occur than not occur. It does not have dimensions, so its range goes from 0 to infinite and in brief it functions as follows: when the OR is equal to 1, it means there is no association; when the OR has a value greater than 1, it means the association is positive (i.e., the presence of the factor is associated with a

more frequent occurrence of the event); and when the OR has a value less than 1, it means the association is negative (i.e., the presence of the factor is associated with a less frequent occurrence of the event). See Figs 1 and 2.

	Muerte	No muerte	Total
DrotAA	210	640	850
Placebo	259	581	840
Total	469	1221	1690

$$OR = \frac{210 \times 581}{640 \times 259} = 0.736$$

OR < 1 → The association is negative (protective factor?)

Fig. 1. OR calculation.

	Muerte	No muerte	Total
DrotAA	210	640	850
Placebo	259	581	840
Total	469	1221	1690

AR of DrotAA =	210 / 850 = 0.247
AR of placebo =	259 / 840 = 0.308
RR =	0.247 / 0.308 = 0.8019 (80.19%)
RR < 1	The association is negative
ARR	0.308 - 0.247 = 0.061 (6.1%)
ARR > 0	The association is negative
RRR =	1 - 0.8019 = 0.198
NNT =	1 / 0.061 = 16.39

Fig. 2. OR calculation. Severe community-acquired pneumonia as a cause of severe sepsis: data from the PROWESS (Recombinant Human Activated Protein C Worldwide Evaluation in Severe Sepsis) study, to evaluate the effect of drotrecogin alfa (activated) (DrotAA) (Laterre *et al.*, 2005).

On the other hand, the risk expresses the likelihood of an adverse result. It is expressed in units that go from 0 to 1 (i.e., with no risk to a risk of 100 %). It requires a period of reference and reflects the accumulated incidence of a disease or event of interest in that period of time. From this emerges the concept of absolute risk (AR), incidence or incidence rate that corresponds to a proportion that can be defined as the number of people who present the event of interest at a certain time (new events) over the number of people at risk at that point. Then, the incidence rate or AR is always calculated for a period of time.

The RR is a quotient between the risk in the group with the study factor and the risk in the reference group. It is a ratio between the risk of a certain event occurring in the group exposed or operated on compared to the control group. It does not have dimensions, so its range goes from 0 to infinite and in brief it functions as follows: when the RR is equal to 1, it means there is no association; when the RR has a value greater than 1, it means the association is positive (i.e., the presence of the study factor is associated with a more frequent occurrence of the event); and when the RR has a value less than 1, it means the association is negative (i.e., the presence of the factor is associated with a less frequent occurrence of the event). See Fig. 3.

	Muerte	No muerte	Total
DrotAA	210	640	850
Placebo	259	581	840
Total	469	1221	1690

AR of DrotAA =	210 / 850 = 0.247
AR of placebo =	259 / 840 = 0.308
RR =	0.247 / 0.308 = 0.8019 (80.19%)
RR < 1	The association is negative
ARR	0.308 - 0.247 = 0.061 (6.1%)
ARR > 0	The association is negative
RRR =	1 - 0.8019 = 0.198
NNT =	1 / 0.061 = 16.39

Fig. 3. RR, ARR and NNT calculation diagram. Severe community-acquired pneumonia as a cause of severe sepsis: data from the PROWESS (Recombinant Human Activated Protein C Worldwide Evaluation in Severe Sepsis) study, to evaluate the effect of drotrecogin alfa (activated) (DrotAA) (Laterre *et al.*, 2005).

The measures of impact on the other hand are the “absolute risk reduction” (ARR), the “relative risk reduction” (RRR), the “number needed to treat” (NNT) and the “number needed to harm” (NNH).

The ARR expresses how much the study intervention reduces the risk compared to the subjects who do not receive it. In other words, this is the difference between the risk in the control group and the risk in the group with the study factor. It has no dimensions and its range goes from -1 to 1 because it calculates by subtracting the RR from the value 1 that represents 100 % of the subjects. In brief it functions as follows: when the ARR is equal to 0, it means there is no association; when the ARR has a value less than 0, it means the association is positive (i.e., the presence of the factor is associated with

a more frequent occurrence of the event); and when the ARR has a value greater than 0, it means the association is negative (i.e., the presence of the factor is associated with a less frequent occurrence of the event).

The RRR, also call the attributable fraction or relative risk difference, is the quotient between the absolute decrease of the risk and the risk of the control group or, which is the same, the difference between the risk of the group in which the experimental therapy or test is applied minus the risk of the control group or standard therapy divided by the risk in the control group. $RRR = RAR / R_c = (R_c - R_e) / R_c$ has the same characteristics as the ARR. However, it has one shortcoming: it does not differentiate the very great risks or benefits from the very small ones and does not vary according to the sample size (Manterola & Otzen, 2015).

The NNT is a term introduced by Laupacis *et al.* (1988). It was proposed in the context of RCT to evaluate the impact of a therapy. It is defined as the number of individuals to treat with the experimental therapy in order to produce, or to avoid, an additional event compared to what would occur with the control therapy. It is easily calculated, since it is the inverse of the ARR, or $1/RAR$ (Manterola, C.; & Otzen, 2015).

The NNH is the opposite of the NNT. This means that a negative NNT indicates that the therapy has a detrimental effect (the experimental therapy is of less benefit than the control or the standard), or that the adverse effects inherent to the therapy are greater in the experimental group. In other words, the NNT represents the number of people needed to try to produce an effect in 1 of them; and the NNH is the number of people needed to try to produce harm in 1 of them. The lower the NNT, the greater the magnitude of the therapy effect at issue. When the NNH is higher, the risk of causing harm with the new therapy or TP is lower. These calculations make it possible to evaluate not only the magnitude of the effects but also the cost-benefit of the intervention. If the NNT of a drug has a value close to the NNH, the possibilities of improving the patient are similar to the possibilities of bringing about some harm; therefore, this drug actually has little chance of being useful. The RCT conducted well includes the NNT and the NNH, or they at least include the data needed to make the calculations. There are calculators online that can easily obtain the NNT and the NNH with their respective 95 % CI (<http://www.calctool.org/CALC/prof/medical/NNT>).

Every article must clearly indicate the statistical tools used in the analysis process of the study, from the

simplest to the most complex. As most of the articles mention descriptive and analytical statistical tools, both to perform bivariate and variable analyses, it is not uncommon to mention the use of tools to assess the magnitude of the effect of the study therapy or TP compared to the standard in use or another one with its respective 95 % CI. In the case of comparative studies, the reporting of such tools is indispensable, or at least that the authors publish the numbers with which a reader can obtain the values of impact and measures of effect.

Thus, the most appropriate ways to represent the results in a clinical trial are the OR, AR, RR, ARR, RRR, NNT and NNH (Laupacis *et al.*; Cook & Sackett, 1995; Sackett *et al.*, 2000), and the statistical significance is nothing more than that, “the statistical significance”, which can sometimes be positive and clinically irrelevant, or negative, without that necessarily meaning that there are real differences between the study variables.

SCORING SYSTEMS AND CHECKLISTS

The following deals with initiatives by different groups that study research methodology, which have contributed different tools to help in the general and specific assessment of the methodological quality of articles.

CONSORT: “Consolidated Standards of Reporting Trials”. This was developed to guide authors to improve the publication quality of randomized CT. It is checklist that consists of 5 domains (Title and summary, introduction, methods, results and discussion) that include 22 items, in which the description of a series of variables inherent to a CT are evaluated. Among the items, it asks authors to create a flow chart to describe the steps of the study participants, from selection and recruitment, distribution of the therapy, follow-up and analysis. It is one of the most commonly used tools and is constantly updated (Moher *et al.*, 2001; Zwarenstein *et al.*).

QUORUM: “Quality of Reporting of Meta-analyses”. It was developed to guide authors to improve the publication quality of meta-analyses. It is a checklist that consists of 5 domains (abstract, introduction, methods, results and discussion of a meta-analysis) organized into 21 categories and subcategories relating to searches, selection of primary articles, evaluation of validity of articles, data extraction, study characteristics, synthesis of the methodological quantitative data, etc. Additionally, a flow chart is required that provides information regarding the CT included and excluded and the reasons for their exclusion (Moher *et al.*, 1999).

STROBE: “Strengthening the Reporting of Observational Studies in Epidemiology”. This was developed to guide authors to improve the publication quality of observational studies. This statement consists of 5 domains (Title and abstract, introduction, methods, results and discussion) that include 22 items, 18 of which are of general application for cohort, case-control and cross-sectional studies, and 4 that are specific to each of the three designs. It also requires that the authors create a flow chart (Vandenbroucke *et al.*).

TREND: This was developed to guide authors to improve the publication quality of studies that use non-randomized designs. This statement has 4 domains (Title, abstract and introduction, methods, results (includes a flow chart of the participants) and discussion) that include 22 items. It is meant to assess a non-randomized CT and its guidelines emphasize the presentation of the theories used, the description of the intervention, the conditions of comparison, the research design used and the methods of adjustment for possible biases in the studies that use non-randomized designs (Des Jarlais *et al.*).

MInCir: “Metodología de Investigación en Cirugía” (English: Research Methodology in Surgery). This was developed to assess the methodological quality of studies with different designs in relation to articles about therapy or TP. It is made up of 3 domains: the first is related to the study design, the second to the size of the study population, and the third to the description of the methodology used in the study (mention of the aims, justification of the design, sample eligibility criteria and justification of the sample). Thus, a final score is generated that can vary between 6 and 36 points, assigning 6 points to the study of lower methodological quality and 36 points to one of better methodological quality (Manterola *et al.*, 2006a; Manterola *et al.*, 2015).

MINORS: “Methodological index for non-randomized studies”. This was developed to guide authors to improve the publication quality of non-randomized studies in the area of surgery, comparative or not. It contains 12 items: the first eight for the non-comparative studies, the remaining items for the comparative studies (Slim *et al.*, 2003).

In conclusion, anyone writing a manuscript related to therapy or TP or reading an article of this type must demand at the very least a clear, precise and concise objective with respect to the research conducted, explicit mention of the design used with the respective inherent methodological details, and the mention and execution of statistical tools and related measures, or at least the numbers needed to calculate these values.

OTZEN, T.; MANTEROLA, C. & MINCIR GROUP. Cómo evaluar e interpretar un artículo científico sobre tratamiento o procedimientos terapéuticos. *Int. J. Morphol.*, 35(2):776-784, 2017.

RESUMEN: Independiente del área temática y diseño empleado, se ha verificado que entre el 40 % y 60 % de los estudios publicados en revistas científicas del ámbito biomédico, corresponden a artículos de tratamiento o procedimientos terapéuticos (PT). Quien escribe un manuscrito relacionado con tratamiento o PT, o quien lee un artículo de este tipo debe exigir al menos un objetivo claro, preciso y conciso respecto del escenario de la investigación que se realizó; la mención explícita del diseño empleado con los respectivos detalles metodológicos inherentes a este; y la mención y ejecución de herramientas estadísticas y medidas de asociación, o al menos los números necesarios para poder calcular estos valores. El objetivo de este manuscrito es presentar una síntesis de los elementos fundamentales para una correcta escritura, lectura y valoración de este tipo de artículos, independiente del área disciplinaria en la que tenga origen la investigación realizada.

PALABRAS CLAVE: Tratamiento; procedimientos terapéuticos; terapéutica; riesgo; incidencia; ensayo clínico; estudios de cohorte; revisión sistemática.

REFERENCIAS BLIOGRÁFICAS

- Caballero-Mateos, A. M.; Martínez-Cara, J. G. & Redondo-Cerezo, E. Hepatic hydatid cysts causing biliary obstruction. *Clin. Gastroenterol. Hepatol.*, pii:S1542-3565(17):30012-5, 2017.
- Cook, R. J. & Sackett, D. L. The number needed to treat: a clinically useful measure of treatment effect. *B. M. J.*, 310(6977):452-4, 1995.
- Dawson, B. D. & Trapp, R. G. *Bioestadística Médica*. 4ª ed. México, El Manual Moderno, 2005.
- Des Jarlais, D. C.; Lyles, C.; Crepaz, N. & TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am. J. Public Health*, 94(3):361-6, 2004.
- Djerf, P.; Montgomery, A.; Hallerbäck, B.; Håkansson, H. O. & Johnsson, F. One- and ten-year outcome of laparoscopic anterior 120° versus total fundoplication: a double-blind, randomized multicenter study. *Surg. Endosc.*, 30(1):168-77, 2016.
- Feinstein, A. R. Clinical epidemiology: The architecture of clinical research. *Stat. Med.*, 14(11):1263, 1995.
- Fisterra.com – Atención Primaria en la Red. Madrid, Elsevier, 2017. Disponible en: <http://www.fisterra.com/guias-clinicas/>
- Fletcher, R. H.; Fletcher, S. W. & Wagner, E. H. *Epidemiología Clínica. Aspectos Fundamentales*. 2ª ed. Barcelona, Masson, 2002.
- Frazzoni, M.; Manno, M.; De Micheli, E. & Savarino, V. Intra-oesophageal acid suppression in complicated gastro-oesophageal reflux disease: esomeprazole versus lansoprazole. *Dig. Liver Dis.*, 38(2):85-90, 2006.
- Grozavu, C.; Ilias, M. & Pantile, D. Multivisceral echinococcosis: concept, diagnosis, management. *Chirurgia (Bucur)*, 109(6):758-68, 2014.
- Hennekens, C. H. & Buring, J. E. *Epidemiology in Medicine*. Boston, Little Brown & Company, 1987.
- Hernández, B. & Velasco, H. E. Encuestas transversales. *Salud Pública Mex.*, 42(5):447-55, 2000.
- Kelsey, J. L.; Whittemore, A. S.; Evans, A. S. & Thompson, W. D. *Methods in Observational Epidemiology*. 2nd ed. New York, Oxford University Press, 1996.

- Laupacis, A.; Sackett, D. L. & Roberts, R. S. An assessment of clinically useful measures of the consequences of treatment. *N. Engl. J. Med.*, 318(26):1728-33, 1988.
- Liu, P.; Yang, J. M.; Niu, W. Y.; Kan, T.; Xie, F.; Li, D. Q.; Wang, Y. & Zhou, Y. M. Prognostic factors in the surgical treatment of caudate lobe hepatocellular carcinoma. *World J. Gastroenterol.*, 16(9):1123-8, 2010.
- Manterola, C. & Grande, L. Methodological quality of articles on therapeutic procedures published in Cirugía Española. Evaluation of the period 2005-2008. *Cir. Esp.*, 87(4):244-50, 2010.
- Manterola, C. & Otzen, T. Bias in clinical research. *Int. J. Morphol.*, 33(3):1156-64, 2015.
- Manterola, C. & Otzen, T. Experimental studies 1st part. Clinical trial. *Int. J. Morphol.*, 33(1):342-9, 2015.
- Manterola, C. & Otzen, T. Observational studies. The most commonly used designs in clinical research. *Int. J. Morphol.*, 32(2):634-45, 2014.
- Manterola, C. Revisión sistemática de la literatura: Síntesis de la evidencia. *Rev. Med. Clin. Condes*, 20(6):897-903, 2009.
- Manterola, C.; Asenjo-Lobos, C. & Otzen, T. Hierarchy of evidence: levels of evidence and grades of recommendation from current use. *Rev. Chil. Infectol.*, 31(6):705-18, 2014.
- Manterola, C.; Busquets, J.; Pascual, M. & Grande, L. What is the methodological quality of articles on therapeutic procedures published in Cirugía Española? *Cir. Esp.*, 79(2):95-100, 2006b.
- Manterola, C.; Cartes-Velasquez, R. & Otzen, T. Instructions for the use of MInCir Scale to assess methodological quality in therapy studies. *Int. J. Morphol.*, 33(4):1463-7, 2015.
- Manterola, C.; Pineda, V. & Grupo MInCir. El valor de "p" y la "significación estadística". Aspectos generales y su valor en la práctica clínica. *Rev. Chil. Cir.*, 60(1):86-9, 2008.
- Manterola, C.; Pineda, V. & Vial, M. Acerca de los ensayos clínicos en el ámbito quirúrgico. Reflexiones, dificultades y desafíos. *Rev. Chil. Cir.*, 61(5):482-9, 2009.
- Manterola, C.; Pineda, V. & Vial, M. Open versus laparoscopic resection in non-complicated colon cancer. A systematic review. *Cir. Esp.*, 78(1):28-33, 2005b.
- Manterola, C.; Pineda, V.; Vial, M. & Grande, L. How should the results of a research study be presented? II. The manuscript and the publication process. *Cir. Esp.*, 81(2):70-7, 2007.
- Manterola, C.; Pineda, V.; Vial, M., Losada, H. & Muñoz, S. Surgery for morbid obesity: selection of operation based on evidence from literature review. *Obes. Surg.*, 15(1):106-13, 2005a.
- Manterola, C.; Pineda, V.; Vial, M.; Losada, H. & MINCIR Group. What is the methodologic quality of human therapy studies in ISI surgical publications? *Ann. Surg.*, 244(5):827-32, 2006a.
- Manterola, D. C. & Otzen, H. T. Clinical Risk Assessment, Interpretation and Practical Utility. *Int. J. Morphol.*, 33(3):842-9, 2015.
- Manterola, D. C. Medicina Basada en la Evidencia o Medicina Basada en Pruebas I. Generalidades acerca de su aplicación en la práctica clínica cotidiana. *Rev. Med. Clin. Condes*, 20(1):125-30, 2009.
- Moher, D.; Cook, D. J.; Eastwood, S.; Olkin, I.; Rennie, D. & Stroup, D. F. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet*, 354(9193):1896-900, 1999.
- Moher, D.; Schulz, K. F.; Altman, D. & CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*, 285(15):1987-91, 2001.
- Riegelman, R. K. *Studying a Study and Testing a Test: How to Read the Medical Evidence*. 6th ed. Philadelphia, Wolters Kluwer / Lippincott Williams & Wilkins, 2013.
- Rossi, M.; Barreca, M.; de Bortoli, N.; Renzi, C.; Santi, S.; Gennai, A.; Bellini, M.; Costa, F.; Conio, M. & Marchi, S. Efficacy of Nissen fundoplication versus medical therapy in the regression of low-grade dysplasia in patients with Barrett esophagus: a prospective study. *Ann. Surg.*, 243(1):58-63, 2006.
- Sackett, D. L.; Richardson, W. S.; Rosenberg, W. & Hynes, R. B. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd ed. London, Churchill-Livingstone, 2000.
- Slim, K.; Nini, E.; Forestier, D.; Kwiatkowski, F.; Panis, Y. & Chipponi, J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. *A. N. Z. J. Surg.*, 73(9):712-6, 2003.
- Vandenbroucke, J. P.; von Elm, E.; Altman, D. G.; Gøtzsche, P. C.; Mulrow, C. D.; Pocock, S. J.; Poole, C.; Schlesselman, J. J.; Egger, M. & STROBE initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Ann. Intern. Med.*, 147(8):W163-94, 2007.
- Zwarenstein, M.; Treweek, S.; Gagnier, J. J.; Altman, D. G.; Tunis, S.; Haynes, B.; Oxman, A. D.; Moher, D.; CONSORT group & Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *B. M. J.*, 337:a2390, 2008.

Correspondence to:
Dr. Carlos Manterola
CEMyQ and Department of Surgery
Universidad de La Frontera.
Manuel Montt 112
office 408
Temuco
Chile

E-mail: carlos.manterola@ufrontera.cl

Received: 27-02-2017
Accepted: 15-03-2017