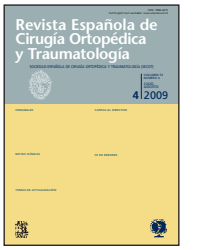
Rev esp. cir. ortop. traumatol. 2009;**53(4)**:261-270

Revista Española de Cirugía Ortopédica y Traumatología

www.elsevier.es/rot



REVIEW ARTICLE

Evidence-based orthopedic and trauma surgery: characteristics and evaluation criteria of research studies

A. Lizaur

Deafting committee, Elda General Hospital, Alicante, Spain

Received October 22, 2008; accepted January 26, 2009
Available on the internet from June 5, 2009

KEYWORDS
Evidence-based medicine;
Types of studies;
Studies in orthopaedic and trauma surgery

Abstract
In response to the new clinical decision-making approach called Evidence-based Medicine, several authors nowadays include a section stating the level of evidence of their study in order to make it easier for readers to select the papers that are most relevant to them on a specific medical topic.
Moreover, and even if this cannot be required of other orthopedic surgery journals to which we usually submit our manuscripts, we think it might be advisable to carry out a self-criticism exercise of our work in order to assess its genuine quality and significance.
To do this, we believe it is extremely important to know the standardized ranking systems and the publication guidelines for different kinds of studies that are commonly used as these can help us apply the most appropriate methodology in each case and make it easier for the reader to assimilate the text, thereby contributing to the enhancement of its scientific quality. This review paper discusses the hierarchies and levels of evidence most frequently employed in the field of Orthopedic and Trauma Surgery, as well as the basic requirements that must be fulfilled by scientific manuscripts.

PALABRAS CLAVE
Medicina basada en la evidencia;
Tipos de estudios;
Estudios en cirugía ortopédica y traumatología

Cirugía ortopédica y traumatología basada en la evidencia: características y criterios de evaluación de los estudios de investigación

Resumen
Como respuesta a la nueva tendencia de toma de decisiones clínicas denominada medicina basada en la evidencia, diversas publicaciones han introducido un apartado en el resumen que identifica el nivel de evidencia del estudio, a fin de facilitar la selección de trabajos a los interesados sobre una cuestión médica determinada.

E-mail: lizaur1@telefonica.net

1888-4415/$ - see front matter

Por otro lado, y aunque no sea exigible en otras revistas de cirugía ortopédica a las que habitualmente enviemos nuestros trabajos, nos parece recomendable realizar una autocrítica para conocer su calidad y la importancia relativa que pudiera derivarse de ellos. Para ello, creemos que es muy importante conocer la jerarquización estandarizada y las guías de publicación de los diversos tipos de estudios, pues estos instrumentos pueden elevar la calidad científica del proyecto, al aplicar la metodología más idónea y facilitarle al lector su asimilación crítica.
En este trabajo de revisión se exponen las jerarquías y los niveles de evidencia de los tipos de estudios más frecuentes en la especialidad de cirugía ortopédica y traumatología, así como las características básicas que cada uno de ellos deben reunir para establecer sus categorías.

Introduction

The first reference to Evidence-based medicine (EBM) was made in an article written and published by clinical epidemiologists Sackett and Guyatt, founders of the Evidence-based Medicine Working Group of the University of McMaster, Ontario, Canada (EBMWG)[1]. This article set the theoretical foundations of the movement regarding a new perspective on decision making in medicine.

EMB consists in making decisions or adopting clinical criteria on the basis of the best evidence provided by studies of the highest quality as well as on one's own experience[2]. The aim is that the best scientific information should be available for physicians to use in their medical practice. It is also true that some authors believe that if, out of fear of making a mistake, physicians were to adopt the suggestions blindly –the conclusions of a meta-analysis, for instance— their clinical freedom would be affected and, in some cases, a reduction of costs would be induced[3,4], though this in itself would not be harmful, since when resources are limited it is our obligation as medical doctors to strive for our work to reach as many patients as possible.

As regards the kind of studies usually carried out within the field of orthopedic and trauma surgery and trauma, it must be admitted that clinical doctors in general find it difficult to understand the bibliography, despite the great number of elementary books on the topic. This is due to the fact that knowledge of epidemiology and statistics is often deficient in pre-graduate courses, and as a result doctors are not familiar with the language used in these fields. The situation has led to the publication in journals of summaries of epidemiologic and statistical techniques ready to be used in the production of publishable material[5].

This review has a twofold objective: 1) to make the understanding of methodological concepts easier for specialists in orthopedics and 2) to divulge the current standards for quality studies.

The foundations of Evidence-based Medicine

EBM requires the consideration of three complementary and synergic aspects[6]:

— one's own personal experience, together with the experience acquired in the professional environment of clinical practice.
— a personal and critical evaluation of the medical literature on a concrete issue: the search for the most relevant publications (in medical databases); a critical reading of these data and the assignment of a level of evidence those publications based on the methodological quality of the studies involved (without taking into account the prestige of the authors or institution where they were carried out).
— the patients' understanding of the procedure and the consideration of their preferences.

Relevant studies

The aim is to find out, with respect to a specific medical topic, which studies contribute the most solid evidence. They tend to be published studies, since these have usually been submitted to processes of critical review and correction by external specialists. At present, there is an excess of publications in medical journals, which makes it impossible to read all that is published and requires a selection to be made, apart from considering that, in some studies, there could be acknowledged and unacknowledged interests on the part of the authors.

It must also be taken into account that the most widespread journals do not necessarily contain the best studies, and that these journals tend to introduce a bias by publishing only the studies with good results, thus leaving out others, especially those with negative results. Moreover, a medical doctor could be using a diagnostic or therapeutic method that has traditionally given him/her good results without having published it, mainly in cases of widely used conventional treatment, this being precisely why he/she makes the choice —and rightly— of using it, although EBM may contemplate other alternative methods or state that they are better.

We must bear in mind that there is a significant difference between the role of publications in biomedical journals and the role of professional recommendations appearing in clinical guidelines. The former constitute repositories of knowledge where authors expose their

research and experience on a specific topic, which may or may not be extrapolated to other situations; the latter are recommendations by groups of specialists, often based on the publications mentioned above, which express an agreement regarding the best possible treatment in the light of the evidence obtained, and are made public as protocols, consensus documents, clinical guidelines, etc.

Hierarchy of tests or evidence level

The evidence level method assesses the genuine quality of the tests presented in a study. It appears to be true that the better designed a research paper is, the greater the reliability of its results and conclusions. In order to be able to quantify the amount of methodological rigor and excellence, we must create a scale that allows us to produce a hierarchy of evidence levels with a view to selecting the most recommendable articles for the clinical topic we are dealing with[7,8].

If our specific aim is to assess the effects of a certain treatment and categorize the evidence provided (therefore leaving aside prognostic and diagnostic studies), we can choose to use one from among a variety of scales that have been designed according to the type of clinical study and the methodology used. The scales are all quite similar, but the most frequently used ones are the scale of the US Agency for Healthcare Research and Quality (AHRQ)[9] (table 1) and, in our professional medium, the scale of the Oxford Centre for Evidence-Based Medicine (CEBM)[10] (table 2). An extension of this scale is the one used by *The Journal of Bone and Joint Surgery (Am)*[11] and by *Clinical Orthopaedics and Related Research*[12]. This scale also includes studies to improve and assess diagnostic capacity and that is adapted to the most frequently published studies in Orthopedic and Trauma Surgery. The latter scale is shown on table 3 (we have not included the columns corresponding to economics or management decision making so as to make it simpler and because they are infrequent in our field). In this scale, a meta-analysis is given the same level of evidence as an individual study and never a higher level.

Table 1 AHRQ evidence level of studies[9]

Evidence level	Type of study
I	Highly recommendable results and evidence
Ia	Meta-analysis of well-designed randomized controlled clinical tests
Ib	Well-designed randomized controlled test
II	Recommendable results and evidence
IIa	Well-designed non-randomized controlled study
IIb	Not wholly experimental well-designed study (cohort study)
III	Recommendable but inconclusive results and evidence
IIIa	Descriptive, non-experimental well-designed studies, such as comparative, correlative or case-control studies
IV	No evidence of research
IVa	Series studies
IVb	Documents or opinions from specialist committees, or clinical experience of prestigious authorities

AHRQ: Agency for Healthcare Reasearch and Quality.

Table 2 CEBM evidence level of studies[10]

Evidence level	Type of study
Ia	Systematic review (meta-analysis), with homogeneity, of well-designed randomized controlled clinical tests[a]
Ib	Well-designed randomized controlled clinical test with narrow confidence interval[b]
Ic	"All or none" type studies[c]
IIa	Meta-analysis (with homogeneity) of cohort studies
IIb	Cohort study or randomized minor quality clinical test[d]
IIc	Outcomes research[e]
IIIa	Meta-analysis (with homogeneity) of case-control studies
IIIb	Case-control study
IV	Case series or cohort studies or minor quality case control study
V	Specialist opinion without explicit critical assessment and not based on physiology

CEBM: Centre for Evidence-Based Medicine.

[a] The level of evidence is not entirely conclusive; however, the revision is a systematic one with statistically significant heterogeneity.

[b] The evidence level in the randomized clinical trials is not considered entirely conclusive; the confidence interval is too wide and not statistically significant.

[c] When all patients die before having availed themselves of a certain treatment and they survive when we they do have the treatment available, or when some patients died before availability of treatment and none die when treatment is within reach.

[d] For example with a wide confidence interval or with 80% follow-up.

[e] Cohort studies in patients with the same diagnosis in which the effects of the treatment received are considered.

[f] Without a clear definition of the groups compared and/or without an objective measurement of exposures and events and/or without appropriately identifying or controlling known confounding variables and/or without a complete follow-up.

Table 3 Evidence levels for research studies, according to journals specialized in orthopedic and trauma surgery[11,12]

	Therapeutic studies	Prognostic studies	Diagnostic studies
	Research into treatment results	Research into the prognosis of a process or condition with a risk factor	Research into a diagnostic method
Level I			
a	High quality clinical test[a]	High quality prospective study[b]	Confirmation of previously described diagnostic method in a successive series of patients (compared with standard)
b	Meta-analysis[c] of level I studies with homogeneous results	Meta-analysis[c] of level I studies	Meta-analysis[c] of level I studies
Level II			
a	Minor quality clinical test[d]	Minor quality prospective study[e]	Development of new diagnostic criteria in a consecutive series of patients (compared with standard)
b	Unrandomized prospective comparative study[f]	Retrospective study	Meta-analysis[c] of level II studies
c	Prospective cohort[g] study	Untreated controls obtained from controlled randomized study	
d	Meta-analysis[c] of level II studies or of level I with inconsistent results	Meta-analysis[c] of level II studies	
Level III			
a	Case-control study[h]	Case-control study[h]	Non-consecutive patients study, or inconsistency when compared with standard
b	Retrospective comparative study		Meta-analysis[c] of level III studies
c	Meta-analysis[c] of level III studies		
Level IV			
a	Prospective or retrospective case series[j]	Case series	Case-control study
b			Studies with no comparison with standard
Level V	Expert opinion[k]	Expert opinion[k]	Expert opinion[k]

[a] Clinical trial: high quality prospective randomized controlled comparative study, with differences that may be statistically significant or not but with a narrow confidence interval.

[b] High quality prospective diagnostic study (all cases were included at the same stage along the course of the disease and with over 80% follow-up of cases.

[c] Metaanalysis: systematic review of at least 2 prior studies.

[d] Clinical trial: lower-quality prospective randomized controlled comparative study, loss of 420% of cases. The study was unmasked, poorly randomized, only analyzed a few aspects related to the results, etc.

[e] Lower-quality prospective diagnostic study: cases included at different stages along the course of the disease or follow-up of less than 80% of cases.

[f] Prospective comparative study: comparison of patients treated in the same institution, some with one method (eg., cemented prosthesis) and the others with a different method (eg., uncemented prosthesis.

[g] Cohort study.

[h] Control cases: patients identified by their outcome (eg., cases, patients with a failed prosthesis; controls, patients with a successful prosthesis).

[ia] Prospective non-randomized comparative study.

[j] Series of cases: prospective or retrospective, treated with one method without making a comparison with other methods or the standard.

Types of study

Glossary

Regarding the different types of study that can be carried out, there is a series of basic well-defined concepts[13] that have not always been understood in our professional medium and that are needed to define the characteristics of the studies correctly. Their English counterparts can be found in the review topics of other foreign orthopedic journals[5].

Retrospective: the patients existed *before* the study was designed and started. Data is collected after the medical procedure takes place. An example would be the review of the results of treatment in patients who have already been treated.

Prospective: patients start being identified and included *after* the study has been designed. Its development takes place over a period of time determined by the research protocol. Data is collected as the medical acts are performed. An instance of this would be to design the study of a treatment first, and then to include the patients who will undergo the treatment.

In spite of the fact that these two terms are widely used in biomedical articles, they are semantically ambiguous, and, because of this, their use is not advised at present. The terms that are preferred are "cohort" or "case-control", depending on whether the data collection is retrospective or prospective, on whether the specification of hypotheses states they are previous to or independent from the data and whether the indication regarding the aim of the study is to evaluate effects (prospective cohort study), to evaluate predictive capacity (prospective cohort) or to search into the causes (prospective cohort and retrospective case-controls).

Controlled: the first step the researcher takes is the design of a research protocol where the following determinations should be included: selection criteria (inclusion and exclusion), for his case, how distribution into groups will be carried out, the variables that will be studied and the evaluation methods. All these actions will prevent the occurrence of biases or confusion variables.

Randomized: the selection of the patients that will make up the study series consists in determining what patients are eligible and deciding on the method by which they will be included in a particular group (a study or control group, in comparative studies). The distribution method should be as random as possible, all the patients having the same chance of being included in a particular group, so that the differences of the factors that could affect the results in the two groups are reduced. The phrase "well randomized" is used to refer to a reasonably valid method in which the randomness sequence is adequately concealed until the assignation is completed. An appropriate and simple way to do this is having an external observer write a list of numbers (that can be produced randomly in a computer) with which the researcher will assign the patient to a certain treatment, following the numerical order (for example, even or uneven numbers—zero being even), and without knowing what treatment each number stands for. Alternatively, the researcher could place the assignation in a closed envelope that will be opened at the moment when the patient is to be given the treatment. The latest trend consists in assigning random assignations by means of a central computer: by internet or by telephone, the researcher provides the patient's initial data and the computer assigns the treatment. When the assignation is random and also considers the patient's characteristics, thus minimizing differences between groups, it is called "dynamic".

There are other ways of randomizing that are considered *not reasonably valid*, such as those performed by birth date, clinical history number, date of appointment, alternate assignation, etc, since it has been shown that these methods may produce a bias in the assignation of the patients.

Masked or blind: this refers to the manner in which the study groups are selected and concealed, and by which their results are analyzed: *a)* non-blind: the researcher, and sometimes the patient, knows what group they belong to (study or control group); *b)* blind: the patient does not know what group he/she belongs to (informed consent is required); *c)* double blind: the medical doctor who carries out the procedure does not know the group, or may know it only at the moment of carrying it out, and *d)* triple blind: when the group is unknown to the patient, the doctor, and the external observer that analyzes and evaluates results.

The ideal solution, in order to attain a higher degree of precision, would be to specify the type of masking or blinding that is used: whether it involves the patient, the person applying the treatment, the person in charge of follow-up, the one who controls the evolution, etc.

Possible biomedical studies

A great variety of studies has been described, each one of them having a specific goal. This is due to the fact that not all studies are equally appropriate for a particular aim[14]. In biomedical research the three main objectives are the therapeutic, the prognostic and the diagnostic objectives.

The aim of therapeutic studies is to assess the effect or result of medical treatment (medication, surgery, etc.), thus they aspire to determine whether it is possible "to change the future"; they require an adequate assessment of the treatment. The most appropriate kind of study for this aim is the clinical test (prospective and comparative with a standard of reference), and, with a much lower evidence level, the case series.

The aim of prognostic studies is to assess the implication of different factors in the appearance of a condition or event (for example, preoperative factors that can cause the incidence of postsurgical infection), and they thus aspire to "predict the future"; they require an appropriate assessment of the degree of success. To attain this objective, a longitudinal observation study (follow-up extends over a period of time) is far more adequate than a clinical test. The most suitable studies, in a decreasing order of evidence level, are cohort and case-control studies.

Finally, the aim of diagnostic studies is to determine the variables that will enable a correct diagnostic classification; they also require an appropriate assessment of the degree of success. The most adequate types of study for this aim are the transverse studies (at one moment in time) in which the reference variable (the standard) and the diagnostic method are assessed simultaneously.

There are many classifications of studies; however, we believe that the most suitable one for the least experienced professional is the following:

— Experimental study: in this kind of study the researcher manipulates exposure to a certain factor (for example, administering a treatment) in a number of individuals that is then compared with another group that was exposed to a different factor or not exposed at all. The most classical and reliable study is the clinical test. An example of this is the work of Marqués et al[15], which set off from the hypothesis that the PFN nail afforded better results than the gamma nail in unstable trochanteric femur fractures. The authors created a scenario by designing a prospective research protocol and selecting patients with a particular condition (unstable trochanteric fracture). Controlled manipulation was performed on these patients, who had been randomly placed in two groups (they were exposed to one of the two surgical treatments). Next, the preoperative features and postoperative results of the two groups were analyzed and compared, with the aim of detecting whether the results obtained with one treatment were superior, equal or inferior to those obtained with the other. In a word, the key point in an experimental study is that it is the researcher who should assign treatment and, if he does so randomly, the study will attain the highest validity.
— Observational or non-experimental study: after exposure or when the experiment is not feasible, a non-experimental study that will somehow simulate the impracticable experiment can be designed. The researcher does not intervene; he/she only observes the variables that were judged to be of interest for the study (for example, in a group of patients who have already undergone treatment, the results and pre-therapeutic factors that could have had an effect on the outcome are evaluated). An observational study may be analytic or descriptive and, depending on the moment it is performed, it may be transverse or longitudinal. An example of this kind of study is the work of Oñorbe et al[16], in which a series of cases with floating knee injuries were analyzed retrospectively. Differently from what is commonly done in experimental studies, the authors of this study did not perform controlled manipulation (treatment had already been administered), and they only observed and assessed the pre- and post-therapeutic interest variables fixed during the design of the study. All studies having a prognostic or diagnostic objective with no researcher-assigned variable, owing to the absence of treatment, can be considered to be observational studies.
— Analytic observational study: statistical analyses are carried out with the aim of establishing association or causality relations between the variables being analyzed (for example, why, how and to what degree certain factors have caused the appearance of an event, a particular outcome, or the evolution of a condition). When an analytic study is decided upon, a great deal is usually known about the condition or event, thus specific hypotheses previously drawn from descriptive studies can be tested. The typical analytic studies are cohort and case-control studies. An example of this kind of study is the work of Pardo-LLopis et al[17], in which pre-op and post-op metal (Co and Cr) concentrations in serum were analyzed prospectively in a group of patients with a metal-metal total hip prosthesis. Serial assessment was carried out up to 4 years following treatment, and it was found that a higher metal concentration in serum was associated with a higher inclination of the acetabular component. These studies are helpful for establishing hypotheses regarding possible treatment and then comparing them with the results of clinical tests. For example, patients can be randomly assigned to groups with different levels of inclination of the acetabular component.
— Descriptive observational study: this kind of study does not establish statistical relations; it only describes the frequency of a problem and its most common characteristics, and identifies the most vulnerable population groups and the risk factors. The classical descriptive studies are the outcomes study, the prevalence (or transverse) study and the case series study. An example of this is the work of Duart et al[18], who designed a retrospective study selecting a group of coxarthrosis patients of less than 50 years of age, and analyzed the previous factors that could have led to the early development of the condition (epidemiologic, clinical and radiologic evaluation). They concluded that more than half the cases presented with signs of anterior femoro-acetabular impingement, a coxarthrosis risk factor in young patients.
— Outcomes study: This kind of study is observational and descriptive, and could be compared to the case series study but it uses extra data about the whole of the population instead of isolated information about an individual. It describes the incidence of the condition in the population in relation with the interest variables (for example, age, tobacco intake, social situation, etc.). The main drawback of this type of study is that it cannot determine whether there exists association between exposure and an individual condition.
— Transverse study: this kind of study is also observational and descriptive. It amounts to a prevalence study that is carried out at a specific moment or period. Its aim is to identify the patients that present with a certain condition or event at a specific moment in time, regardless of how long they will have the condition or when it started (for example, all the cases with hemophilic arthropathy being treated at that moment, not including possible future patients). This study analyzes risk factor exposure and the condition or event simultaneously, and therefore does not reveal the temporal sequence of the events, thus making it impossible to determine whether exposure to one factor preceded or came after the condition. An example of this type of study is the work of García-Bógalo et al[19], who identified the cases, in a period of 7 years, that had undergone treatment with anterior intramedullary nailing for a diaphyseal fracture of the humerus. The purpose was to find out whether the treatment produced an injury of the rotator cuff (estimated with an ultrasound scan) and the risk factor when using an anterior approach. It was found that

although 17% of the patients presented with a cuff injury, this had no statistical relation with poorer functional results. As has already been stated, diagnostic capacity studies are usually transverse.

— Longitudinal study: in this kind of study, the patients are identified and then submitted to follow-up over a specific period of time, variables being assessed both at the beginning and at the end of the study (for example, a prospective study in which patients are progressively included and evaluated after a previously established minimum period of time.) A case in point is the work of Sanz-Reig et al[20], in which a selection was made of patients of more than 70 years of age implanted with total hip prosthesis due to coxarthrosis within a period of 5 years. The crucial preoperative factors were analyzed and follow-up was established for 7 years after treatment, at the end of which the significant factors and the evolution of the process were evaluated.

Most frequent studies in orthopedic surgery and trauma

The most common studies in orthopedic publications are: 1) cohort studies, case-control studies and case series studies, among observational studies; 2) the clinical test, among experimental studies; and 3) meta-analysis, among evidence review studies.

Guidelines for the production of quality reports

A study may be publishable and of a high quality regardless of its design. The requirements are that it should be well-developed and based on a correct selection, taking into account the aim pursued in the research, as has already been stated above[13,14].

Various entities and groups of experts have written reporting guidelines that are published by the most prestigious orthopedic journals. These guidelines make recommendations for producing works of the highest quality about the most common types of study. They are not strict rules, but rather guiding principles regarding the essential features that a work must have in order to be acceptable. These guidelines do not intend to improve the quality of publications; their aim is to make works easier to understand both for readers as well as for reviewers and editors. Most of these guides are of free access and also easily accessible on the internet, their Spanish version being available as well[21].

Among the best known entities is STROBE (www.strobe-statement.org), which provides recommendations for improving publication guides on three kinds of observational studies (cohort, case-control and transverse studies) in the form of a list of 22 items (STROBE declaration) related to the different sections that make up an original scientific publication[22]. Another is the Consort Group (www.consort-statement.org), which presents the key features that will afford quality to a publication of randomized clinical tests. In addition, Equator (www.equator-network.org) provides publication guides that assist authors, reviewers and editors in spreading the circulation of their work. Lastly, The QUORUM Document (now renamed Prism) consists of a check-list that provides recommendations for producing a meta-analysis[23].

Cohort study

The cohort study is observational, analytic, longitudinal and prospective. The individuals that are selected are all free from the condition or event under study (for example, infection following prosthetic hip surgery) and a follow-up period is established to check if they develop the condition. Patients are distributed according to whether they have, or not, one or various risk factors (for example, obesity) and undergo a follow-up period (each of the patients' periods are estimated) in which the frequency of appearance of the condition or event being studied is observed (post-op infection). If, at the end of the observation period, the incidence of the condition (the infection) is significantly higher in the risk factor group (obesity), we can safely conclude that there is a relation between exposure to the risk factor (obesity) and the incidence of the condition (infection).

As can be seen, the aim of the study is to look into the causality between a risk factor and the condition or event being investigated. The hypothetical association can be estimated with the relative risk (RR) variable. In our example it would require the comparison, by means of a 2?2 table, of the frequency of occurrence of the infection in the risk factor group (obesity) against that of the non-risk group. The mathematical calculus would involve the ratio between the incidence in the exposed group (the number of infections in the obese group divided by the total number of obese patients) and the incidence in the unexposed one (the number of infections in the non-obese group divided by the total number of non-obese patients). The resulting ratio must be complemented with a confidence interval.

This type of study is the ideal tool, even better than clinical tests, for prognostic research. It offers the advantage of enabling the estimate of incidence and relative risk and the possibility of studying several results for each exposure factor. The drawbacks are its inefficiency in uncommon events or in cases with long latency periods, and that it may require large series. The recommendations of the STROBE Group are very helpful for improving the quality of the presentation of the experience.

An example of this is the work of García-Álvarez et al[24], in which one of the objectives was to assess the incidence of an infection in a specific context. With this aim in mind, a prospective study was designed selecting a cohort of elderly patients with subcapital fracture who had been treated with total hip replacement. The whole group was free from illness (infection), and some risk factors were determined before surgery (immunologic and nutritional variables, estimated by means of serum analysis). They were given follow-up treatment, and it was observed that the group with the lowest concentration of IgE had a greater probability of infection of the surgical wound than the group with normal or high values of IgE. This opens the door for new investigations (clinical tests, ideally) in which the incidence of infection may be reduced by treating the IgE.

Case-control study

This is an observational, longitudinal, analytic and retrospective study in which patients with a condition or medical event of interest (for example, rigidity after knee prosthesis implant) are selected, and then compared with other individuals (controls) without the condition or event (for example, with no rigidity after knee prosthesis implant). Next, both groups' antecedents are examined for the presence of one or more risk factors or for exposure to past risk factors that could have had an effect on the development of the condition or event being studied (for example, under 100° preoperative knee flexion).

The degree of risk is estimated by comparing the frequency of occurrence of the condition (knee rigidity) in the risk factor group (under 100° preoperative flexion) with the non-risk group (above 100°). As the incidence of the condition is not available, it is not possible to work out the relative risk. In its place an odds ratio is used with a mathematical calculus, by means of a 2?2 table, corresponding to the quotient between the probability that the event will happen (the exposed number of cases multiplied by the number of unexposed controls; in our example: number of cases with rigidity and under 100° previous flexion ? number of cases without rigidity and over 100° previous flexion) and the probability that it will not happen (number of exposed controls multiplied by number of unexposed cases; in our example: number of cases without rigidity and under 100° previous flexion ? number of cases with rigidity and above 100° previous flexion).

The control group can be formed by individuals without the condition but with a risk of acquiring it, and assigning one or more of them to each one of the cases according to pair-matching criteria that must be fixed a priori in the design of the work. It would be ideal if these selection criteria were of the type that would result in each control being as similar as possible to its corresponding case, so that the only difference is the presence of the condition being studied. In this way the only differences or similarities that can appear between the cases and their respective controls concern the risk factors to be analyzed.

This type of study is apt for etiologic research. It was designed for the evaluation of the risk factor; however, it is possible to apply it in the study of treatment exploring the possible origin of collateral and adverse effects, or complications after treatment whose low frequency of occurrence makes a clinical test impracticable. It must be remembered, however, that this type of study has a weak validity for establishing the causal relation between the adverse effect and treatment, since no randomizing or masking is used.

Case-control studies have the advantage of being useful for analyzing uncommon events, in relatively small series; they evaluate many risk factors for one condition or event and provide odds ratio estimation. The recommendations of the STROBE Group are very helpful for improving the quality of reports on the experience.

An example of this kind of study is the work of García-Tejero et al[25], a retrospective study in which a series of elderly women with wrist fracture was selected (cases) and compared with another group of healthy elderly women (same sex, similar age, but without a wrist fracture). Both groups were submitted to a densitometry of the calcaneus; a significantly greater mineral bone density was observed in the control group. This led the researchers to conclude that osteopenia is a risk factor for wrist fracture in elderly women, and therefore the estimation of osteopenia could predict the appearance of the fracture.

Case series study

The case series study is a descriptive kind of study that consists in the simple identification and observation of a group of clinical cases that have appeared over a certain period of time, and is also used to evaluate a therapeutic model within a specific time period.

The disadvantage of this type of study is that it has no control group as reference (the standard) or for comparison, and no randomizing or blinding either. This affords the study a much lower evidence level than that of the clinical test, and its results must be compared against those obtained by other authors with the consequent difficulty that arises regarding the selection of patients and the feasibility of making a valid comparison with groups in other studies. It is worth pointing out, however, that although these studies are the weakest form of clinical research, they can sometimes provide the only practical or available information for supporting a therapeutic strategy, especially in the case of infrequent processes or when the evolution of treatment takes place before the production of randomized study designs in the medical practice. It may also be the only practical design for radically different treatment, for example, amputation versus surgery to save a limb.

One example is the retrospective study carried out by Torres-Torres et al[26], in which a series of fractures of the radial head are shown and the results after treatment with bipolar prostheses are presented.

Clinical test

This is the prototypic experimental study and the standard one among therapeutic studies. In general, and particularly when drugs are being studied, clinical tests must be sanctioned by a bioethics committee after they are designed and before the study is begun. Also, the patients involved must know the aims of the study, its risks and benefits, and they must sign an informed consent.

The clinical test is a longitudinal, non-observational, prospective study. It begins with the presentation of a hypothesis drawn from uncontrolled, descriptive or retrospective, observational studies, or from preclinical studies. It ends when the terms defined in the protocol reach their end, or prematurely, when the effects are clearly harmful or beneficial to the patients involved.

These are the steps of the basic method: *a)* planning and protocolization of the study; *b)* selection of method and patients; *c)* description of the method of random patient assignation with a specific group, and of the means of concealing the randomizing sequences; *d)* description of the basic characteristics of the groups, confirming their homogeneity, and indication of loss to follow-up and its causes in each group; *e)* information about the type and

form of the actions taken with each group; *f)* data collection of the variables of the study, according to previously designed protocol, also in the cases of masking; and *g)* interpretation and analysis of results with adequate statistical techniques and according to the established confidence level, also in the cases of masking.

Thus, well-designed clinical tests must fulfill the requirements mentioned above: they must be prospective, controlled, comparative and random. As regards masking, it is not always feasible in a clinical test, especially in the case of certain kinds of treatment, such as surgery or radiotherapy. At times, the study fails after the assignation of treatment because procedures or adverse effects vary considerably between groups.

With respect to the size of the series, it must be calculated probabilistically in order to prevent the possibility of not finding differences between the groups when there actually *are* differences (type II error). The main solution we have for reducing random errors is to enlarge the size of the series, thus increasing accuracy and reducing confidence intervals and standard error. There are alternative solutions to increase the accuracy of the study, such as the selection of a more reliable response variable (lower inter-case variability) and the adjustment of the analysis according to initial conditions.

A clinical test is considered to be high quality when, apart from being well-designed, the losses to follow-up of the series are under 20%, and the differences between groups are statistically significant or, if this is not so, confidence intervals are small.

A clinical test is considered to be low quality when it is not well-designed, especially if due to wrong randomizing or low quality of the masking, when the results are based on less than 80% of the subjects initially recruited, or when the confidence intervals for the differences between groups are too wide.

There are various methods for evaluating the quality of a clinical test, such as the recommendations of the Consort Group mentioned above and the contrastive Jadad scale[27], one of the most highly used methods due to its simplicity. This scale gives one point to each one of the following issues: study is described as randomized; study is described as double blind; losses to follow-up, as well as withdrawals or drop-outs are described. One additional point is given to each one of the following: randomizing method is reported and valid; and blinding method is reported and valid. Thus, the total scoring varies between 0 and 5; a score of 3 or less is considered low quality; a score of 4 stands for quality; and a score of 5 is considered to be excellent quality.

An example of this kind of study is the work of Vicent-Vera et al[28], in which a group of patients was prospectively selected and treated with knee arthroscopy. The patients were distributed into 3 groups by means of a list of random numbers. At post-op they were injected with an intra-articular analgesic (two of them with different doses of morphine and one with a placebo); intensity of pain and demand of adjunctive analgesia were then estimated. The 3 groups showed similar characteristics. It was observed that the morphine injection was effective as postsurgical analgesia in this kind of treatment; however, there were no significant differences between the doses.

Experimental tests with drugs

With a view to the commercialization of drugs, the Spanish Agency of Drugs and Health Products[29] established the following as the basic phases of a study:

- Phase I studies: These are initial safety studies that adjust an initial dose and evaluate its adverse systemic effects in a group of highly selected healthy patients. These are called efficacy studies.
- Phase II studies: These are safety studies carried out on individuals presenting with the condition being studied (a specific condition). Their aim is to adjust doses and estimate dose/response. They are also considered to be efficacy studies.
- Phase III studies: They are formed by the IIIa and IIIb phase studies. The IIIa phase studies are effectiveness studies and are carried out on a large population group presenting with the condition being studied; the effectiveness of the treatment is compared with a control treatment. These are often multi-center studies. Dose/response, clinical effectiveness and collateral effects of the medication being studied are evaluated. Results are sent to the health authorities for evaluation before being approved for commercialization. Phase IIIb studies are effectiveness studies whose aim is to obtain the additional information or carry out the extra research that the authorities deem necessary before the medication is approved for commercialization.
- Phase IV studies: These are effectiveness and efficacy studies (cost-effective studies) that are carried out after commercialization by different research groups. Their aim is to estimate effectiveness in different population groups (such as special patient groups) and to compare the drug with other drugs whose effectiveness has been tested. These constitute the majority of the studies that are published.

Meta-analysis

Meta-analysis is the quantitative analysis of two or more independent studies with the aim of integrating their findings, describing the characteristics of their results and the degree of homogeneity of the observations. The information obtained from different publications on a specific problem, variable or procedure is submitted to a systematic and structured review (generally computerized medical bibliography databases). Next, the studies whose design fulfills the previously mentioned selection criteria are identified and reviewed, with the intention of obtaining evidence for the resulting larger series from these smaller ones.

The two major methodological difficulties of the meta-analysis of clinical tests are: *a)* the heterogeneity of the studies included, in terms of clinical and socio-demographic characteristics of the populations in each test, the classifications of individuals or interest variables, the therapeutic methods and their indications, the clinical evaluation methods used, etc.; and *b)* the possible biases resulting from publication due to the fact that not all clinical tests with negative or unexpected results are published. The recommendations made by the QUORUM

Group[23] are very helpful to improve the quality of the presentation of these works.

Conclusions

The criteria derived from EBM have a practical application, both for the general practitioner interested in a specific topic as well as for those who wish to evaluate the quality of the scientific works they read or write.

A hierarchical categorization of the various types of research study has been established, as well as the methodological requirements they must fulfill in order to be categorized and so that their evidence level (or strength of collected proof) can be determined.

The understanding of these categories and requirements is an instrument that we can use to carry out a self-criticism of our work. This will help us improve our productions to the greatest of our possibilities and afford them a major scientific relevance.

Conflict of interests

The authors declare they have no economic or personal conflict of interests.

References

1. Evidence-Based Medicine Working Group Evidence-based medicine: a new approach to teaching the practice of Medicine. JAMA. 1992;268:2420-5.
2. Elstein AS. On the origins and development of evidence-based medicine and medical decision making. Inflamm Res. 2004;53:184-9.
3. Grahame-Smith D. Evidence-based medicine: Socratic dissent. BMJ. 1995;310:1126-7.
4. Evidence-based medicine, in its place [editorial]. Lancet. 1995;346:785.
5. Petrie A. Statistics in orthopaedic papers. J Bone Joint Surg Br. 2006;88B:1121-36.
6. Haynes RB, Sackett DL, Guyatt GH, Tugwell P. Clinical epidemiology: how to do clinical practice research. 3.a ed. Philadelphia: Lippincott, Williams and Wilkins; 2005.
7. Jovell AJ, Navarro-Rubio MD. Evaluación de la evidencia científica. Med Clin (Barc). 1995;105:740-3.
8. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature: IX. A method for grading health care recommendations. JAMA. 1995;274:1800-4.
9. Evidence-base practice. US Agency for Healthcare Research and Quality [citado 10 Sep 2008]. Disponible en: http://www.ahrq.gov.
10. Levels of evidence and grades of recommendation. Oxford: Centre for evidence-based medicine [citado 10 Sep 2008]. Disponible en: http://www.cebm.net.
11. Instructions to authors. Levels of evidence for primary research questions. The Journal of Bone and Joint Surgery, American volumen [citado 10 Sep 2008]. Disponible en: http://www.ejbjs.org.
12. Levels of evidence for primary research questions. Clinical Orthopaedics and Related Research [citado 10 Sep 2008]. Disponible en: http://www.editorialmanager.com/corr/account/LOE.doc.
13. Cobo E, González JA, Muñoz P, editors. Bioestadística para no estadísticos: principios para interpretar un estudio clínico. Barcelona: Elsevier; 2007.
14. Fernández E, Cobo E, Guallar-Castillón P, editores. La declaración STROBE o cómo mejorar la presentación de los estudios observacionales. Gac Sanit. 2008;22:87-9.
15. Marqués F, León A, Mestre C, Ballester J, Cáceres E. Estudio prospectivo comparativo entre el clavo gamma trocantérico y el clavo femoral proximal AO (PFN) en el tratamiento de las fracturas trocantéricas inestables. Rev Ortop Traumatol. 2005;49:11-6.
16. Oñorbe F, Ferrer-Santacreu EM, Rodríguez-Merchán EC. Rodilla flotante: revisión retrospectiva de 15 casos tratados durante un periodo de 5 años. Rev Esp Cir Ortop Traumatol. 2008;52:283-9.
17. Pardo-LLopis JV, Martínez-Algarra JC, Sendra-Miralles F, Palomares-Talens E. Determinación de los niveles en suero de cobalto y cromo en 17 pacientes tras el implante de una prótesis total de cadera con par metal-metal. Rev Esp Cir Ortop Traumatol. 2008;52:77-83.
18. Duart J, Del Río J, Mönckeberg J, Valentí JR. Atrapamiento fémoro-acetabular anterior: signos radiológicos en pacientes jóvenes diagnosticados de coxartrosis. Rev Esp Cir Ortop Traumatol. 2008;52:71-6.
19. García-Bógalo R, Larrainzar-Garijo R, Díez-Nicolás E, Llanos-Alcázar LF. Valoración clínico-radiográfica de la agresión al manguito rotador en el abordaje anterógrado del enclavado humeral. Rev Esp Cir Ortop Traumatol. 2008;52:2-8.
20. Sanz-Reig J, Lizaur-Utrilla A, Plazaola-Gutiérrez J, Cebrián Gómez R. La artroplastia total de cadera porosa en pacientes mayores de 70 años. Rev Esp Cir Ortop Traumatol. 2007;51:198-204.
21. Cobo E, editor. Listas de comprobación para autores, revisores y editores de revistas médicas. Med Clin (Barc). 2005;125(Supl 1).
22. Von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, en nombre de la Iniciativa STROBE. Declaración de la Iniciativa STROBE (Strengthening the Reporting of Observational studies in Epidemiology): directrices para la comunicación de estudios observacionales. Gac Sanit. 2008;22:146-52.
23. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet. 1999;354:1896-900.
24. García-Álvarez F, Al-Ghanem R, García-Álvarez I, Larrad R, González-Murga P, López-Baisson A, et al. Estado inmunológico de ancianos candidatos a artroplastia tras fractura subcapital de cadera. Rev Ortop Traumatol. 2007;51:309-13.
25. García-Tejero PM, García-Frasquet MA, Borrás A, Carpintero R, Nieto I. La densitometría de calcáneo permite discriminar fracturas de muñeca en mujeres ancianas. Rev Ortop Traumatol. 2007;51:214-7.
26. Torres-Torres M, Ortiz-Espada A, Corella-Montoya F, Vaquero-Martín J. Estudio de 15 prótesis de cabeza radial bipolares. Rev Esp Cir Ortop Traumatol. 2008;52:306-14.
27. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et-al Assessing the quality of reports on randomized clinical trials: Is blinding necessary?. Controlled Clin Trials. 1996;17:1-12.
28. Vicent-Vera J, Palomo JM, Díaz-Almodovar JL. Morfina intraarticular tras artroscopia de rodilla: estudio prospectivo, aleatorizado y doble ciego. Rev Ortop Traumatol. 2007;52:188-93.
29. Ensayos clínicos. Agencia Española del Medicamento y Productos Sanitarios [citado 5 Nov 2008]. Disponible en: http://www.agemed.es/profHumana/ensayosClinicos.