

# The interpretation of the results of a randomised controlled trial

**IDOIA GAMINDE**

TEACHING AND EDUCATION SERVICES, INVESTIGATION AND HEALTH DEVELOPMENT.  
HEALTH DEPARTMENT. GOVERNMENT OF NAVARRE, SPAIN

**JUAN ERVITI**

DRUG PRESCRIBING SERVICE. NAVARRE REGIONAL HEALTH SERVICE IN SPAIN

**The aim of this article is to revise some of the key concepts related to randomised controlled trials used to demonstrate the efficacy of pharmacological treatments: the role of bias, the role of chance, clinical relevance, the external validity of trials, surrogate and composite endpoints, analysis of secondary endpoints and subgroups, non-inferiority trials, the influence of individual studies in meta-analysis and finally, the opinion of the patient.**

**02**

The interpretation of the results of a randomised controlled trial

**12**

Duration of treatment with citicoline in moderate and severe ischemic stroke

**15**

Notice: early suspension of the ACCORD trial

## The interpretation of the results of a randomised controlled trial

In order to correctly interpret a randomised controlled trial (RCT) firstly, an evaluation of the strength of the results must be made: is the benefit of the intervention due to the method employed in the study?

### Bias

Bias refers to any systematic error due to design, conduct, or interpretation of a study. The most habitual bias are those errors made during random allocation and blinding.

Randomisation consists of randomly allocating each participant to one of the groups of intervention. This is done to ensure that the groups included in the trial are homogenous in all relevant characteristics, except for the intervention that each receive<sup>1</sup>. The absence of random allocation (or errors in the process of allocation) can result in creating groups not comparable to each other. In fact, in many systematic reviews, non-randomised trials are excluded because of the bias that can occur. It is important that the sequence of random allocation of patients to each group remains unknown to both the patients and the investigators.

Blinding is a process which ensures that neither the subjects participating in a RCT, nor the observers, or both, are aware of the intervention assigned to the subjects, or of any other characteristic that may cause a bias in the results<sup>1</sup>.

Studies with incorrect blinding can overestimate effects of the intervention up to 41% while studies in which blinding is unclear, up to 30%<sup>2</sup>.

### The role of chance

Once checked for bias related to the method of randomisation and blinding, we examine whether the results obtained after the trial are real or are due to chance.

The comparison of efficacy of two interventions can be carried out by either verifying hypothesis (using tests of statistical significance) or by tests of estimation (confidence intervals).

Tests of statistical significance calculate the probability that the results obtained between the groups of the RCT are due to chance supposing that interventions were equally efficacious (null hypothesis). This probability constitutes the level of statistical significance and is represented by the letter “p”. Generally the value  $p=0.05$  is adopted as a cut off point below which the hypothesis that both interventions possess an equal effect (with a confidence interval of 95%) can be rejected and the result can be concluded *statistically significant*<sup>3</sup>. This only allows us to either reject or accept the null hypothesis that there are “no differences” between the two groups, but does not give any indication of the magnitude of the difference or of its direction.

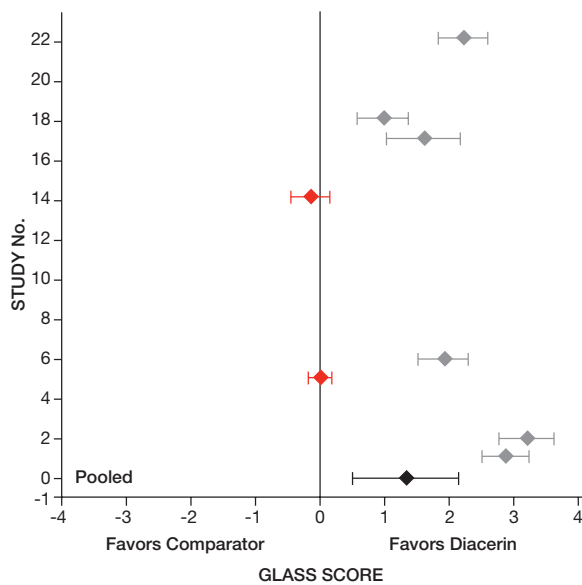
Confidence intervals (CI) are most useful in evaluating the role of chance. Confidence intervals give an idea of the magnitude or relevance of an observed effect. They allow us to discover between what limits do the real differences probably lie. Habitually, a CI of 95% is employed, which means that in 95% of the times, a correctly constructed CI should contain the true value of a variable in question. For example, if an observed difference between two treatments is 22% ( $p<0.05$ ; 95% CI, 17 to 27%) this means, that there is 95% confidence that the real difference between the two treatments lies between 17 and 27%.

A useful characteristic of CI is that it allows investigators to state whether statistical significance has been reached or not, just like a hypothesis test. When the measure of an effect is the **difference** between interventions, if the CI contains the zero value, then the result is considered not statistically significant. This means that no differences have been shown in effects of treatment in the intervention group with regard to the control group. When, instead of an absolute difference, our interest is in relative differences like **relative risk**, if the CI includes the value 1, then the result is considered not statistically significant<sup>3</sup>.

For example in figure 1 we observe different CI's in the difference between diacerein and placebo in pain relief. Two studies cross the line of “no differences” (the CI contains the value 0), while six other studies show superiority of dia-

Randomisation and blinding are key aspects of a RCT

**Figure 1.** A meta-analysis of the efficacy of diacerein vs placebo in the treatment of pain in osteoarthritis.



Source: [1] Rintelen B, Neumann K, Leeb B. A meta-analysis of controlled clinical studies with diacerein in the treatment of osteoarthritis. Arch Intern Med. 2006;166(17):1899-906.

cein against placebo. We can also observe a difference in range of the CI, that reflects the precision of the estimations, where narrower intervals are more precise than wider ones.

Another advantage of CI on the verification of hypotheses is that they offer additional information. The upper and lower limits of CI inform us about how small or large the real effect can reach. If the CI is narrow, then we can be confident that any effect outside the interval is ruled out from the study. This situation presents when the size of the sample is very large and the estimate of the real effect is very precise. This means that the study has sufficient “power” to detect an effect.

But if the study is small, and the CI is very wide, then the interval will capture a diverse number in the extent of the effects. The estimations on size of the effect would thus be quite imprecise meaning that the study has little “power” and offers less information.

### Possible errors in the interpretation of results

Just like the values of “p”, confidence intervals help us interpret the findings of a study in the light of the effects of chance. However, in the process of interpretation of results, there are a few pitfalls:

An error can be made on seeing effects that are not real. The CI show us that the differences are “statistically significant,” and therefore, we conclude that both treatments are different. However, just because it is very unlikely to observe a great difference only by chance, this does not mean that the effect is true. By definition, one in every 20 (5%) significant results could be false and the differences found would be attributed to chance. For this reason, chance can cause us to err on making us believe that there exist differences between groups when in reality they do not exist. This is known as a type I error. The probability of making an error of this type is referred to as “ $\alpha$ ” and is normally expressed as the level of significance “p”. A value of  $p=0.05$  makes reference to the existence of  $\alpha=0.05$ .

Another possible error is to conclude from a non significant result that there is no effect, when in reality one does exist. This is a type II error. Assuming “non significance” with absence of effect is a frequent and harmful misunderstanding. An insignificant CI simply indicates that the difference observed is consistent with the statement that no real differences between the 2 groups exist. But, we cannot reject this possibility. Just because we have not found any effect with the treatment does not mean there is none. The probability of committing this error is usually denoted by the Greek letter  $\beta$  and its complement,  $1-\beta$ , is what is referred to as statistical power or statistical potential.

Thus, when designing a study, the minimum magnitude of the difference or association that will be considered clinically relevant must be established. Likewise, the desired statistical power for the study should be established beforehand and accordingly the size of the sample should also be calculated.

### Statistical significance and clinical relevance

Statistical significance is sometimes interpreted incorrectly by associating it with an important result. Tests of significance only question whether

*Even with well designed RCT's a false positive is created for every 20 results simply due to chance*

*It is necessary to determine the clinical relevance of a statistically significant trial*

the data obtained from a study can be attributed to chance or not. Rejection of equivalence between the two interventions does not necessarily mean that we accept that there is an important difference between them. A large study can find that a small difference is statistically significant. Thus, it is not the same thing to evaluate the clinical relevance of the difference found. In the evaluation of the importance of statistically significant results, the size of the effect (not the size of the significance) is relevant.

To demonstrate clinical relevance of the results of a RCT, the recommended form is to include relative risk reduction (RRR), absolute risk reduction (ARR) and the number necessary to treat (NNT) when presenting the results. For example, 15% of the patients in an intervention group die and 20% in the control group. The relative risk (RR), defined as the ratio between those exposed to the new treatment or activity and those not exposed, is in this case 0.75 (ie:  $0.15/0.20$ ). The mortality risk in patients who receive the new treatment compared to the control group is 0.75. The RRR is the complement of RR, that is,  $(1-0.75)*100=25\%$ . The new treatment reduces the risk of dying by 25% compared to what occurs with the control group. The ARR would be  $0.20-0.15=0.05$  (5%). We can say that for every 100 patients treated with the new therapy, 5 deaths can be avoided. The next question would be: how many patients should we treat to avoid one death? In other words, what is the NNT? Its calculation requires a simple rule of three which is resolved by dividing  $1/ARR$ . In his case  $1/0.05=20$ . Therefore, the solution is that we need to treat 20 patients with the new treatment in order to avoid one death<sup>4</sup>.

This way of presenting results quantifies the efforts made to reduce an unfavourable event. To present results only as a percentage relative risk reduction, though technically correct, tends to magnify the effect of the intervention by describing similarly very unlike situations. Small changes

in the basal absolute risk of an infrequent clinical event can lead to great changes in the number necessary to determine the clinical relevance of a randomised trial that presents statistically significant results. The smaller the NNT, the greater is the magnitude of the effect of the intervention. If no efficacy was obtained by the treatment, the absolute risk reduction would have been zero and the NNT infinite. As occurs with the estimations of other parameters, NNT should be expressed with confidence intervals to estimate the uncertainty that the given parameter presents<sup>4</sup>.

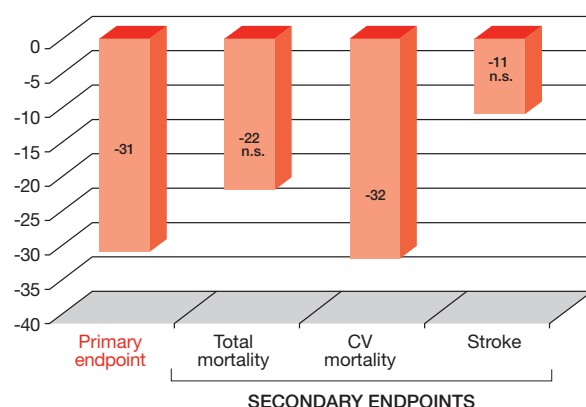
### Resources to calculate the clinical relevance of results

- Online calculator that determines the clinical relevance of RCTs with the corresponding CI: <http://www.healthcare.ubc.ca/calc/clinsig.html>
- Lists of NNTs [<http://www.jr2.ox.ac.uk/bandolier/band50/b50-8.html>].

To illustrate this, different views can be perceived with regards to the efficacy of statins in the primary prevention of cardiovascular events, if we either focus our attention on relative risk and absolute risk, or simply, if we evaluate the fact that there are significant differences between study groups. Of the trials realised, the WOSCOP<sup>5</sup> found the greatest differences in the primary endpoint (non fatal infarction + coronary mortality). If we look at the percentage of relative risk reduction in the distinct endpoints, the efficacy of the intervention is apparently notable (figure 2).

However, if we consider the absolute risk reduction, we find that the effect size is quite poor, despite whether differences were statistically significant or not (figure 3). In the placebo group, some of the events included in the composite endpoint occurred in 7.9% of the patients. On the other hand, in the intervention group, 5.5% of the patients developed some event. That is 2.4% of the patients obtained some benefit, while the remaining 97.6% did not benefit after 5 years of treatment (figure 4). If expressed in terms

**Figure 2.** Relative risk reductions with pravastatin vs placebo. WOSCOP trial.



of NNT, then it would be necessary to treat 42 patients for 5 years in order to avoid one event included in the primary endpoint.

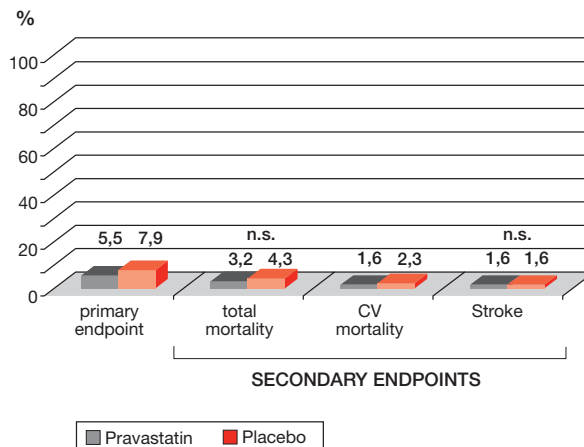
In the rest of the clinical trials published in cardiovascular primary prevention the results were even worse with regard to the effect size of the medication, despite whether the results were statistically significant or not.

There are many other cases where the magnitude of the effect is very small (of debatable clinical relevance), despite the fact that these studies show statistically significant results.

For example, the efficacy of anticholinergic agents in urinary incontinence. In one trial versus placebo, solifenacin showed a reduction in the number of daily micturitions of up to an average of 11 to 10. Even without taking into account the adverse effects of the agent, could the reduction of one micturition a day with respect to a total of 11 micturitions per day be considered clinically relevant?<sup>6</sup>

Apart from this, it is worth mentioning the case of studies that measure the effects of medications by use of scales. Here the differences in scores obtained between the intervention and control groups are evaluated. When compared with relevant endpoints (i.e: myocardial infarction), the determination of clinical relevance of either an improvement or a deterioration of a given magnitude on a scale is objectively more difficult to achieve.

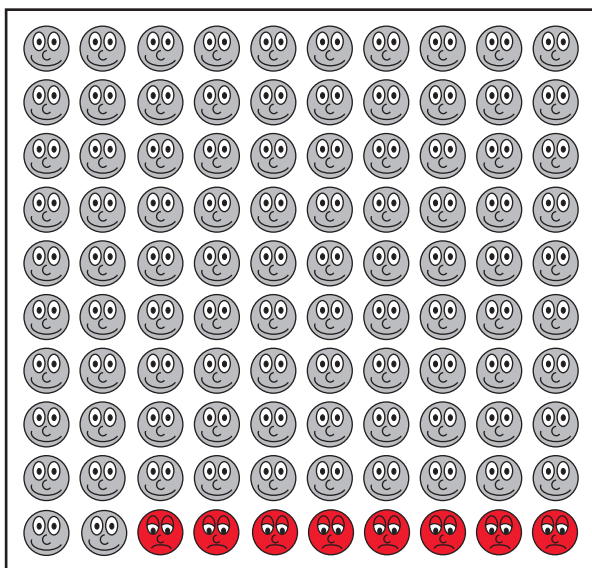
Figure 3. Absolute risk values for different endpoints. WOSCOP trial.



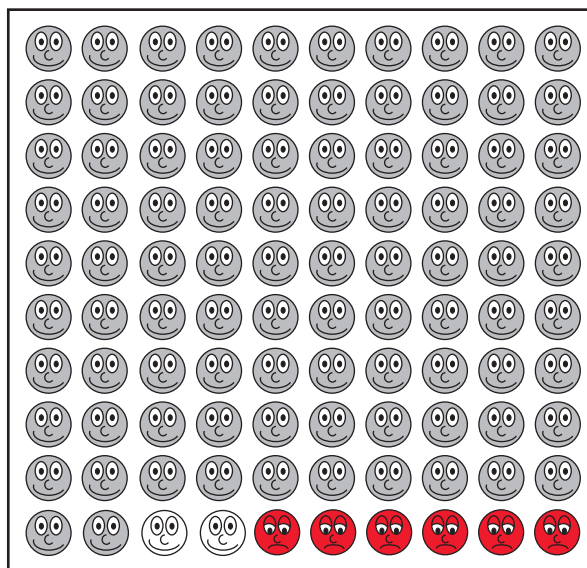
In this sense, a meta-analysis that evaluated the effects of rivastigmin, donepezil, galantamin and memantine on vascular dementia observed an improvement between 1 and 2 points on a 70-point scale (ADAS-cog) in patients treated with these agents. Differences were statistically significant, but the authors themselves questioned the clinical relevance of this finding<sup>7</sup>.

Figure 4. Visual representation of the outcomes in the primary endpoint after 5 years of follow up. WOSCOP trial.

Results on the placebo group after 5 years of follow up



Results on the pravastatin group after 5 years of follow up



- No event
- Either infarction or coronary death prevented
- Infarction or coronary death

*Reflection on the characteristics of patients should be made in order to judge the external validity of the data*

### External validity. To generalise beyond the clinical trial

Another question to bear in mind is that the results of a given study refer only to the patients that have participated in it. Even if an effect is considered probably real and clinically relevant, this still raises another question: can the results obtained be applicable to other groups of patients or to another given patient? Neither the CI nor the p value can aid us in resolving this query. The evaluation of external validity is carried out based on the characteristics of the patients (inclusion and exclusion criteria, proportion of compliance with treatment, etc.) and the setting at which the trial has been developed.

### Other relevant methodological questions

#### The measurement of surrogate endpoints

It is not infrequent to find the use of surrogate endpoints instead of final results with true clinical importance. A surrogate endpoint is a physiological or laboratory finding employed to substitute a hard endpoint that measures how a patient feels, how the patient functions or how long the patient lives. Unless the surrogate endpoints are known to be related to relevant clinical results, they should be interpreted with caution.

#### Examples of surrogate endpoints and the hard endpoints they pretend to substitute

- Blood pressure as an intermediate measure of stroke
- The degree of atherosclerosis in coronary angiography as a measure of myocardial infarction or coronary death.
- Return of spontaneous circulation after cardiorespiratory arrest as a measure of return of neurological function or survival.
- Bone density as a measure of risk of fracture.
- Lipid profile as a measure of cardiovascular risk.
- Glycosilated haemoglobin (HbA1c) as a measure of diabetic control (HbA1c).
- PSA as a marker of prostate cancer.

Ideally decisions should be taken on endpoints such as health-related quality of life, morbidity (infarction or stroke) or even mortality. When surrogate endpoints are used to carry out inferences on given expected benefits, it is assumed that there is a true relationship between the intermediate measurements and the results, that there is a clear connection between the change in values of the surrogate endpoints and a clinically relevant result. More so, when surrogate endpoints give reliable information on relevant clinical results for patients, the effect of the measurement has to be large, strong and of sufficient duration to be able to make inferences on its credibility<sup>8</sup>.

For example, hormone replacement therapy has been employed for many years in postmenopausal women under the conviction that it could provide cardiovascular benefits because of its improvement on the lipid profile. However, when the Women's Health Initiative<sup>9</sup> was carried out the incidence of stroke and coronary disease observed was superior in women who undertook the therapy, despite the better lipid profile. The surrogate endpoint (lipid profile) was not a good predictor of the results of the hard endpoints (stroke or coronary disease).

### Composite endpoints

Composite endpoints are those where two or more endpoints are combined and considered as a single measurement of results. Habitually, they are justified under the presupposition that the effect of each of the components is similar and that patients attribute the same importance to each of the components. However this is not always like this. To correctly interpret composite endpoints, Montori and colleagues<sup>10</sup> suggest the clinician asks the following questions:

- Are the individual endpoints included in the composite endpoint of equal importance to the patients?
- Was there a similar number of events registered among the more or less important endpoints?
- Is it possible that the individual endpoints have similar risk reduction?
- Is the clinical relevance of the individual endpoints similar?
- Among the individual endpoints, are the estimates of the reductions in risk similar and are the CI's sufficiently narrow?

The answers to these questions will determine whether it is necessary to examine the individual endpoints separately.

An example of an inadequate composite endpoint is that employed in the MIRACL<sup>11</sup> trial in which atorvastatin was compared versus placebo in acute coronary syndrome. The proposed endpoint was the sum of the following: death+ non fatal infarction+ resuscitated cardiac arrest + recurrent symptomatic myocardial ischaemia with objective evidence and emergency rehospitalization. On responding to the questions outlined above, the following comments can be made:

One cannot say that all the components have the same importance for patients. For example, “death” or “infarction” or “readmission due to ischaemia” are not comparable in meaning for patients.

The frequency with which different phenomena included in the same composite endpoint occur was very different. The highest incidence observed was readmission to hospital (7.3%) whilst cardiac arrest was only 0.5%.

As for risk reduction observed in each of the individual endpoints, in the hard endpoints (infarction or death) no statistically significant differences were found between the study groups [RR = 0,92 (0,75-1,13)]. However, readmission to hospital showed a greater risk reduction [RR = 0,74 (0,57-0,95)], which ultimately contributed to the outcome on the composite endpoint by showing differences at the limit of statistical significance [RR = 0,84 (0,70-1,00)].

Apart from these considerations, the composite endpoint chosen is not simple, but rather “odd” and it does not seem logical to employ it in evaluating the efficacy of an intervention on the prediction of coronary events.

On the other hand, besides reasoning over whether the chosen endpoint is adequate to the disease under study, it is always important to pay attention to what medications are evaluated in order to make a solid judgement of the composite endpoint. What could occur is that even though a composite endpoint is composed of all important endpoints, the inclusion or exclusion of some of them could condition the final results of the trial.

*Results of composite endpoints should be interpreted with caution*

### Analysis of secondary endpoints

The primary endpoint is that which allows investigators to respond to the objectives of the study. The sample size is calculated in such a way that it includes a sufficient number of individuals that can provide reliable information in the results of that primary endpoint. On the other hand, all trials describe other endpoints that could be interesting, but may not possess a sufficient number of cases to establish solid conclusions. These are secondary endpoints. The proliferation of secondary endpoints in clinical trials can produce results with statistical significance in some of these endpoints solely by chance.

A clear example is the ELITE study where the efficacy of an angiotensin II receptor blocker (losartan) was compared to an angiotensin-converting enzyme inhibitor (captopril) in patients with heart failure. The main composite endpoint was death and/or admission to hospital due to heart failure. No significant differences were found between the two groups. However, a statistically significant difference was found in reduction of total mortality in the case of losartan with respect to captopril. The fact that the investigators found significant differences in the secondary endpoint led them to believe that losartan was superior to captopril in the reduction of mortality.

**Table 1.** Main outcomes in the ELITE trial.

Primary endpoint	Losartan (n=352)	Captopril (n=370)	Risk reduction (CI)	p
Death and/or hospital admission for heart failure	33 (9.4%)	49 (13.2%)	0.32 (-0.04 to 0.55)	0.075
Other endpoints				
Total mortality (secondary endpoint)	17 (4.8%)	32 (8.7%)	0.46 (0.05-0.69)	0.035

**Table 2.** Main outcomes in the ELITE II trial.

Primary endpoint	Losartan (n=1,578)	Captopril (n=1,574)	Hazards ratio (CI)	p
Total mortality (primary endpoint)	280 (17.7%)	250 (15.9%)	1,13 (0.95-1.35)	0.16
Other endpoints				
Sudden death (secondary endpoint)	130 (8.2%)	101 (6.4%)	1,30 (1.00-1.69)	

*The results of secondary endpoints and post hoc analysis only serve to generate hypothesis, and NEVER to take clinical decisions*

For this reason, another clinical trial was designed (ELITE II) in which the primary endpoint was total mortality, and in which similar characteristics of the former study were reproduced. However, in this second trial, no significant differences were found in mortality, and thus the hypothesis generated from the ELITE study was rejected. On the other hand, a reduction in sudden death was observed (secondary endpoint) in 30% of the captopril group with respect to losartan. It would be equally erroneous to consider that losartan offers some advantage over captopril, as to say that captopril has better results in preventing sudden death<sup>12,13</sup> (tables 1 and 2).

In one trial, we can only obtain reliable information of the results concerning the primary endpoint. The data of the secondary endpoints serve to generate new hypotheses that would have to be investigated in future trials.

### Analysis of subgroups.

If a health problem varies in terms of different characteristics, then it would be practical to plan the estimation of the parameter in the different subgroups of interest. If an analysis of the subgroups is desired, then attention should be given to the sample size and the methods of selection of patients. If not carried out this way, then the estimation of the parameter in each subgroup will lose precision in relation to that obtained when analysing the entire sample, since the total number of subjects is clearly inferior. On interpreting the results of subgroups 3 important aspects should be taken into account:

- The definition of the subgroup and the analysis of their results should be planned prior to the development of the trial. If not, then we should be cautious when interpreting the results.
- It is important to evaluate the clinical relevance of the differences found between the various subgroups.

- Statistical management should be realised correctly. Sometimes authors simply publish differences obtained in the primary endpoint in each of the subgroups and then compare them between the same subgroups. This is not a correct approach to interpreting results. The fact that the differences are significant in one subgroup and not in the rest of the subgroups does not mean that there is a real difference between them. The differences found could be conditioned by the different sample sizes of the subgroups or by other motives. A test of interaction should be carried out between subgroups in order to conclude with minimum confidence whether there exists any difference between them<sup>14</sup>.

### Post hoc analysis

*Post hoc* analysis refers to the analysis of a subgroup, not originally defined in the protocol of the study. Normally selected patients are those in which the intervention was effective and they are grouped to find some common characteristic.

An example is the TROPOS<sup>15</sup> study which compared the efficacy of Strontium Ranelate with placebo in the prevention of hip fracture. The results did not show any efficacy of the agent [RR = 0,85 (0,61-1,19)]. A *post hoc* analysis was carried out and differences at the limits of statistical significance were observed in the results of women with high fracture risk (mean age= 80 years, previous fracture among 60% of them and with bone density < -3.5 SD) [RR = 0,64 (0,41-1,00)].

The results of the *post hoc* analysis are not valid to demonstrate or reject the hypothesis of the study and only serve to generate other hypotheses which would need to be investigated further. In no case should we condition our clinical practice by the data obtained by *post hoc* analysis. These analytical strategies serve to generate hypotheses which require verification by adequately designed trials. Another possible use is to generate safety warnings in data bases for adverse drug effects surveillance.

### Non-inferiority trials

Recently, therapeutic equivalence trials and non-inferiority trials are proliferating. The former attempt to show that two interventions are similar from a clinical point of view. To do so, the magnitude of the maximum clinical difference is defined arbitrarily by the investigator who determines what would be acceptable as therapeutic equivalence. This concept is known as delta ( $\delta$ ).



Thus equivalence trials attempt to demonstrate that the effects of an agent under study are found within a range “ $\pm\delta$ ” when compared to the control group.

Non-inferiority trials position themselves on one side of the range of the interval, so as to verify whether the agent under study is below the  $-\delta$  value with respect to the control.

Equivalence trials have been widely employed to evaluate new medications, but they have lost favour to non-inferiority trials. These trials are accepted by regulating authorities to approve new medications or new indications. The use of these trials implies the intention of not trying to show any advantage of the new agent in question with regard to the control agent. An example of this is the COMPASS<sup>16</sup> trial in which the thrombolytic agent, saruplase showed equivalence to streptoquinase in post-infarction, even though an increase in mortality by 50% was observed. Some consider that non-inferiority trials are not ethical as they expose patients to clinical experiments with no security that medication under investigation is not worse than the standard treatment, and without really studying how much better it could be<sup>17</sup>.

### Meta-analysis. Influence of the quality of individual studies.

A meta-analysis provides aggregate information of the results of different individual trials. For this reason, it is important to consider all the questions mentioned up to now for each of the included trials. The differences in quality of these can significantly condition the final outcome of the meta-analysis.

A graphical example of this can be seen in a recently published meta-analysis on the efficacy of diacerein versus placebo in the treatment of osteoarthritis. In this study 19 trials were included to evaluate the results of this agent in the treatment of pain and function. The conclusion reached by the authors was that diacerein provided beneficial effects in both parameters, even though of small magnitude. However, two of the trials were superior in quality in comparison to the rest and in both no significant differences were found between diacerein and placebo. If these 2 trials are compared to the rest, notable differences are observed in the mean scores of the JADAD<sup>19</sup> scale (4.5 versus 2.7), in

*The patient should be offered clear information in terms of absolute risk in order to decide more freely*

the mean duration of the studies (24 versus 2.8 months) and in the mean number of patients included in each trial (404 versus 119). Figure 1.

If only the trials with higher quality were considered in the meta-analysis then the conclusions would have been contrary to those announced by the authors.

### And the patients... what opinion do they have about the treatment we propose?

Today it is widely accepted that patients should actively participate in the decision concerning treatment. The patients' opinion is particularly important in those cases in which the treatment only offers marginal benefits, as in the case of chronic disease. However, medical practice is generally based on the application of clinical guidelines that normally do not contemplate the opinion of the patient.

Patients are not often adequately informed about the treatments prescribed. Usually guidelines are based on recommendations from clinical trials on a large scale that present results in a complex manner. This fact, added to the short time available for both doctors and their patients in the consultancy, makes effective communication difficult between the two parties. However, some studies show that when the efficacy of treatment is explained to patients with emphasis on absolute risk reduction, the majority of them decline the recommended treatment from the guidelines.

The opinion of the well informed patient can foment a more critical approach to clinical investigation and medicine based on clinical guidelines<sup>20</sup>.

## Conclusions

Special attention should be given to random allocation of patients in the study groups and to blinding so as to avoid the main bias of design in a RCT.

Independent of whether a trial has been well designed, one in every 20 findings will produce a statistically significant result, which occurs simply by chance and which in reality is not significant.

Statistical significance does not necessarily mean that the finding is clinically relevant. It is the effect size that determines the importance, not the presence of statistical significance.

It is preferable to employ confidence intervals rather than p values. Both express statistical significance but the CI offers additional information such as the effect size and the precision of the result (amplitude of the interval).

It is important to pay close attention to the characteristics of the patients included in the trial in order to evaluate the external validity of the results obtained.

Results obtained in composite endpoints should be analysed with caution and criteria.

The results obtained from secondary endpoints and post hoc analysis serve to generate hypotheses that should be verified in adequately designed trials. Information from such endpoints should not be translated to clinical practice.

Non-inferiority trials attempt to prove a medication under investigation is not worse than the control treatment, assuming that the variability in the results between both treatments is clinically irrelevant, which is established arbitrarily by the researchers.

The results obtained from a single clinical trial does not justify change in clinical practice. More evidence would be necessary to support or reject the findings of the trial.

Information on the results of clinical trials given to patients should be objective and comprehensible, so that the decision to accept treatment may be made more freely.

### Secondary Journals / Journal Clubs

These carry out a summary with a critique on RCT's and help and facilitate critical analysis

ACP: <http://www.acpjg.org/>

Evidence Based Medicine  
[<http://ebm.bmjournals.com/>]

Evidence Based Medicine en castellano  
[<http://ebm.isciii.es/sumarios.asp>]

Evidence Based Mental Health  
[<http://ebmh.bmjournals.com/>]

Evidence Based Mental Health en castellano  
[<http://ebmh.isciii.es/>]

Evidence Based Nursing [<http://ebn.bmjournals.com/>]

Evidencias en Pediatría  
[<http://www.aepap.org/EvidPediatr/index.htm>]

### Resources on the web on evidence based medicine

<http://www.fisterra.com>

<http://www.infodoctor.org/rafabravo/>

<http://www.redcaspe.org/>

## REFERENCES

1. Laporte J. Principios Básicos de Investigación Clínica. Barcelona: AstraZeneca 2001
2. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273(5):408-12.
3. Argimon J. El intervalo de confianza: algo más que un valor de significación estadística. *Medicina Clínica*. 2002;118(10):382-4.
4. Pita S, López de Ullibarrí I. Número necesario de pacientes a tratar para reducir un evento CAD ATEN PRIMARIA 1998 [cited; 96-8]. Available from: <http://www.fisterra.com/mbe/investiga/5nnt/5nnt.asp>
5. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, Macfarlane PW, et al. Prevention of Coronary Heart Disease with Pravastatin in Men with Hypercholesterolemia. *N Engl J Med* 1995;333(20):1301-8.
6. Chapple CR, Rechberger T, Al-Shukri S, Meffan P, Everaert K, Huang M, et al. Randomized, double-blind placebo- and tolterodine-controlled trial of the once-daily antimuscarinic agent solifenacin in patients with symptomatic overactive bladder. *BJU International* 2004;93(3):303-10.
7. Kavirajan H and Schneider LS. Efficacy and adverse effects of cholinesterase inhibitors and memantine in vascular dementia: a meta-analysis of randomised controlled trials. *Lancet Neurol* 2007;6:782-92.
8. Freemantle N, Calvert M. Composite and surrogate outcomes in randomised controlled trials. *BMJ* 2007;334(7597):756-7.
9. Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *JAMA* 2002;288(3):321-333.
10. Montori VM, Permyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *BMJ* 2005;330(7491):594-6.
11. Schwartz GG, Olsson AG, Ezekowitz MD, Ganz P, Oliver MF, Waters D, et al. Effects of Atorvastatin on Early Recurrent Ischemic Events in Acute Coronary Syndromes: The MIRACL Study: A Randomized Controlled Trial. *JAMA* 2001;285(13):1711-8.
12. Pitt B, Poole-Wilson PA, Segal R, Martinez FA, Dickstein K, Camm AJ, et al. Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial - the Losartan Heart Failure Survival Study ELITE II. *Lancet* 2000;355: 1582-7
13. Pitt B, Segal R, Martinez FA, Meurers G, Cowley AJ, Thomas I, et al. Randomised trial of losartan versus captopril in patients over 65 with heart failure (Evaluation of Losartan in the Elderly Study, ELITE). *Lancet* 1997;349:747-52.
14. Fletcher J. Subgroup analyses: how to avoid being misled. *BMJ* 2007;335:96-7.
15. Reginster JY, Seeman E, De Vernejoul MC, Adami S, Compston J, Phenekos C, et al. Strontium Ranelate Reduces the Risk of Nonvertebral Fractures in Postmenopausal Women with Osteoporosis: Treatment of Peripheral Osteoporosis (TROPOS) Study. *J Clin Endocrinol Metab* 2005;90(5):2816-22
16. Tebbe U, Michels R, Adgey J, Boland J, Caspi A, Charbonnier B, et al. Randomized, double-blind study comparing saruplase with streptokinase therapy in acute myocardial infarction: the COMPASS Equivalence Trial. Comparison Trial of Saruplase and Streptokinase (COMPASS) Investigators. *J Am Coll Cardiol* 1998;31(3): 494-6
17. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007;370:1875-7.
18. Rintelen B, Neumann K, Leeb B. A meta-analysis of controlled clinical studies with diacerein in the treatment of osteoarthritis. *Arch Intern Med* 2006;166(17):1899-906.
19. Jadad A, Moore R, Carroll D, Jenkinson C, Reynolds D, Gavaghan D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clin Trials* 1996;17:1-12.
20. Penston J. Patient's preferences shed light on the murky world of guideline-based medicine. *Journal of Evaluation in Clinical Practice* 2007;13:154-9.

# Duration of citicoline treatment in moderate to severe ischemic stroke

**ANA AZPARREN** DRUG PRESCRIBING SERVICE. NAVARRE REGIONAL HEALTH SERVICE IN SPAIN

**JAIME GÁLLEGO** NEUROLOGY DEPARTMENT. HOSPITAL OF NAVARRA. NAVARRE REGIONAL HEALTH SERVICE IN SPAIN

Citicoline is an intermediary in the biosynthesis of phosphatidylcholine, a component of the cell membrane. During ischemia, phosphatidylcholine breaks down fatty acids that release free radicals, which in turn, augment the ischemic process.

The administration of exogenous citicoline reduces the concentration of free fatty acids, improves neurological signs, preserves levels of phosphatidylcholine and improves neuronal survival.

## Indications

According to the drug information leaflet<sup>1</sup>, the indications of citicoline in ampoules include:

- Treatment of cognitive and neurological disorders associated with the acute and sub-acute (recovery) phase of stroke.
- Treatment of cognitive and neurological disorders associated with head injury.

Recommended doses are between 500-2000 mg /day depending on the severity of the condition.

**The drug information leaflet do not specify duration of treatment.**

Acute treatment refers to initiation of treatment within the first 24-48 hours of the event. Treatment in the sub-acute phase refers to administration of the agent between day 3 and day 14 of the event.

The ampoules can be administered by intramuscular injection, slow intravenous infusion or drop intravenous perfusion.

If necessary the ampoules can be administered orally. The contents should be taken directly or dissolved in half a glass of water.

## Evaluation of efficacy

Three randomised double-blind controlled trials comparing citicoline vs placebo<sup>2,3,4</sup> and a meta-analysis that evaluates 4 previous studies<sup>5</sup> have been published.

The primary endpoint of the meta-analysis was to determine the efficacy of oral citicoline in the recovery of patients with acute moderate or severe ischemic stroke after 12 weeks of treatment, in comparison with placebo. The method employed was the Generalised estimating equations (GEE) recommended by the National Institute of Neurological Disorders and Stroke (NINDS) for trials involving stroke. The GEE represents the probability (OR) of overall recovery of the patient and consists of a combination of the 3 most widely used scales in the clinical evaluation of stroke:

**Barthel index:** evaluates the post-stroke functional disability (feeding, bathing oneself, bowel control), etc.

**National Institute of Health Stroke Scale (NIHSS):** evaluates the neurological state (level of consciousness, orientation, paralysis, etc).

**Modified Rankin scale (MRS):** evaluates function. Values of this scale run from 0 (normal) to 6 (death).

A single scale offers partial information on clinical recovery of the patient, while the value of the GEE allows for a better evaluation of the efficacy of treatment. Overall recovery is defined as NIHSS  $\leq$  1, MRE  $\leq$  1 and BI  $\geq$  95. A GEE value defined as effective does not mean that efficacy has been achieved on all 3 scales.

The secondary endpoints of the meta-analysis were mortality and the evaluation of each of the scales individually.

A criteria for inclusion was a NIHSS score of  $\geq$ 8, which corresponded with moderate to severe stroke. Random allocation was performed in 1,652 patients in the USA, of which 1,372 complied with the criteria of inclusion. The doses employed were 500, 1,000 and 2,000 mg administered for 6 weeks.

#### Primary endpoint: overall recovery after 12 weeks

The lack of efficacy in the group taking 1,000 mg was attributed to the scarce number of participants in the group in addition to greater severity after stroke.

### Mortality

There were no differences in mortality: 18.8% in the citicoline group and 17.8% in the placebo group.

### Safety

The frequency of adverse effects was compared in both groups. In the citicoline group, there was a greater proportion of anxiety and lower extremity oedema, while depression, falls and incontinence was superior among the placebo group. Dose related adverse effects were not specified.

The conclusion of the authors was that in patients with moderate to severe stroke, a daily dose of 2,000 mg of citicoline, 24 hours after the event for 6 weeks increased the probability of overall recovery after 3 months, with no effect on mortality. In any case the results are only modest. Any therapeutic intervention used in patients with mild stroke (NIHSS $<$ 8) would offer little benefit, given that these patients already have a favourable prognosis.

### Conclusion

Based on the meta-analysis outlined above, oral citicoline is recommended for 6 weeks in patients with moderate to severe stroke, both in the acute and sub-acute phase.

The normal dose in the treatment of moderate or severe stroke is 2,000 mg daily.

### Information on consumption in Navarre

In 2007 the total number of patients under treatment of citicoline in ampoules was 1,609. Approximately 56% of them were taking citicoline for more than 2 months, which was excessive. In the figure 1, duration of treatment is shown per patient.

**Table 1.** Overall recovery and mortality after 12 weeks.

#### Probability of overall recovery after 12 weeks of treatment

	Citicoline %	Placebo %	OR (95%CI)	NNT
Any dosis	25.2	20.2	1.33 (1.10 – 1.62)	20
500 mg versus placebo	20.8	15.7	n.s.	
1.000 mg versus placebo	9.1	10.7	n.s.	
2.000 mg versus placebo	27.9	21.9	1.38 (1.1 – 1.72)	17

#### Mortality

	Citicoline %	Placebo %	OR (95%CI)	NNT
	18.8	17.8	n.s.	

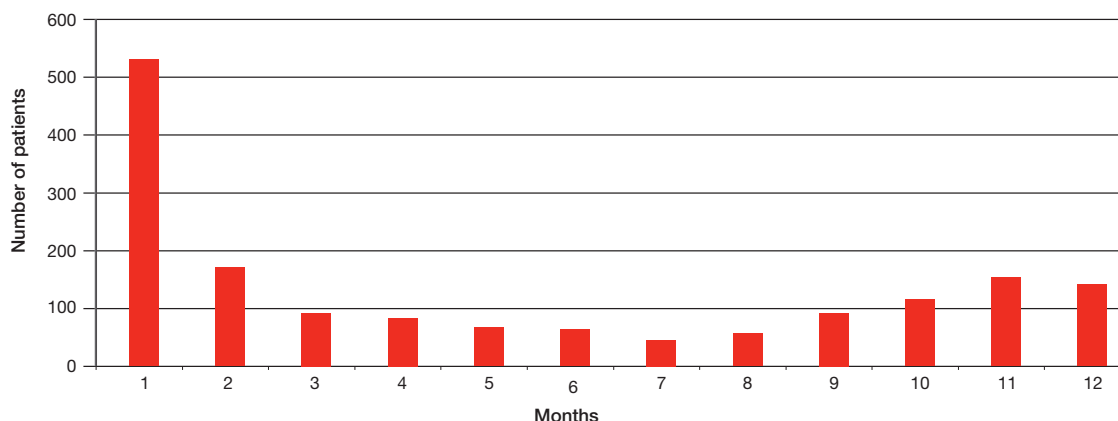
## Future

To confirm the results of this retrospective analysis, in November 2006 another trial was initiated, the ICTUS Study (International Citicoline Trial on acUte Stroke). It is a randomised, multicenter, double-blind and placebo-controlled trial. The objective is to evaluate the effects of a daily dose of 2,000 mg of citicoline in the recovery of the pa-

tients with moderate or severe stroke in the acute phase. There are 2,600 patients. The effects of treatment will be evaluated after 12 weeks and completion of the study is expected in 2010.

This study is financed by the pharmaceutical company Grupo Ferrer S.A. and the main investigator is A. Dávalos.

**Figure 1.** Citicoline treatment duration per patient (Year 2007).



## REFERENCES

1. Somazina© ampollas. Product Information. Ferrer Internacional S.A. Laboratories
2. Clark WM, Warach SJ, Pettigrew LC, Gammans RE, and Sabounjian LA. for the Citicoline Stroke Study Group. *Neurology* 1997;49:671-8.
3. Clark WM, Williams BJ, Selzer KA, Zweifler RM, Sabounjian LA, and Gammans RE. A randomized efficacy trial of citicoline in patients with acute ischemic stroke. *Stroke* 1999;30(12):2592-7.
4. Clark WM, Wechsler LR, Sabounjian LA, and Schwiderski UE for the Citicoline Stroke Study Group. A phase III randomized efficacy trial of 2000 mg citicoline in acute ischemic stroke patients. *Neurology* 2001;57(9):1595-602.
5. Warach S, Pettigrew LC, Dashe JF, Pullicino P, Lefkowitz DM, Sabounjian L, Harnett K, Schwiderski U, and Gammans R. Effects of citicoline on ischemic lesions as measured by diffusion-weighted magnetic resonance imaging. *Ann Neurol* 2000; 48(5):713-22.
6. Dávalos A et al. Oral citicoline in acute ischemic stroke. An individual patient data pooling analysis of clinical trials. *Stroke* 2002;33:2850-7.
7. Protocol 06PRT/3005: ICTUS study: Internacional Citicoline Trial on acUte Stroke (NCT00331890). Disponible en: [www.thelancet.com/journals/lancet/misc/protocol/06PRT-30005](http://www.thelancet.com/journals/lancet/misc/protocol/06PRT-30005). Consultado diciembre 2007.

# Early suspension of ACCORD trial

## EDITORIAL BOARD, BIT

In the last issue of the BIT, mention was made of the ACCORD trial that was expected to shed more light on the efficacy of intensive therapy in type 2 diabetic patients. This consisted of a 2x2 factorial design trial in which a total of 10,251 patients participated and which attempted to respond to the following questions:

- Does an intensive strategy to reduce HbA1c levels below 6% show a greater reduction in the incidence of cardiovascular events than a standard strategy that targets levels of HbA1c between 7.0 and 7.9%?
- In the context of an adequate glycemic control, does a combination of fibrates, (to raise HDL-c and reduce triglycerides), and a statin reduce the incidence of cardiovascular events more than statins alone?
- In the context of adequate glycemic control, does lowering the values of systolic blood pressure to below 120 mmHg reduce the incidence of cardiovascular events to a greater extent than with values below 140 mmHg?

The arm of the trial that attempted to respond to the first question was stopped after 4 years of follow up. The study had a planned duration of between 4 to 8 years, with a mean follow up of 5.6 ye-

**Table 1.** Deaths in Intensive vs Standard Glycemic Control Groups

Standard Glycemic Control	Intensive therapy
203 (11/1,000 persons-year)	257 (14/1,000 persons-year)

ars. The motive for suspension of the trial was an increase in mortality in the arm of intensive therapy with respect to the standard treatment. At the moment the trial was stopped, mean values of HbA1c were 6.4% in patients with intensive therapy and 7.5% in patients in the standard treatment arm.

Complete publication of the results of the suspended arm of the trial is still expected.

## REFERENCES

National Heart, Lung and blood Institute. National Institutes of Health. Action to Control Cardiovascular Risk in Diabetes (ACCORD) Trial.

[http://www.nhlbi.nih.gov/health/prof/heart/other/accord/q\\_a.htm](http://www.nhlbi.nih.gov/health/prof/heart/other/accord/q_a.htm) (last consultation, 27 February 2008)

### Acknowledgements

We thank Dr Clint Jean Louis, of the Emergency Department of the Hospital de Navarra, for translating the original manuscript into English.



**Servicio Navarro de Salud**  
Osasunbidea



#### ISSN

1138-1043

#### COPYRIGHT

NA-1263/1997

#### INFORMATION AND SUBSCRIPTION

Servicio Navarro de Salud / Osasunbidea

Plaza de la Paz, s/n

31002 Pamplona

T +34 848429047

F +34 848429010

#### E-mail

farmacia.atprimaria@cfnavarra.es

#### WEB PAGE

<http://www.cfnavarra.es/WebGN/SOU/publicac/BJ/sumario.htm>

#### EDITORIAL BOARD

Isabel Martín Montaner (president)

Cristina Agudo Pascual

M<sup>ª</sup> José Ariz Arnedo

Jesús Berjón Reyero

José Ignacio Elejalde Guerra

Idoia Gaminde Inda

Maite Hermoso de Mendoza

Rodolfo Montoya Barquet

Lourdes Muruzábal Sitges

Mercedes Noceda Urarte

Javier Lafita Tejedor

Cristina Ibarrola Guillén

Juan Erviti López (coordinator)