Those Were Evidences!

Two months ago I was invited to a meeting to celebrate 30 years of the Cardiovascular Emergency Committee of SAC, and was asked to review the evolution of ideas about evidences in these three decades.

I finished my Residency in 1981 and due to a series of coincidences I ended up being the youngest founding member of the Council in 1982, escorted by a very prestigious group: Raúl Oliveri, Hernán Doval, Marcelo Lapuente, Branco Mautner, Otero y Garzón and Daniel Fernández Bergez. It was a very intense learning period, with fruitful discussions in which we shared practical experiences in the CCU and recently published articles, as well as monthly meetings with debates on complex clinical cases. At the same time, my work as Head of CCU at the Hospital Argerich gave me the opportunity of enjoying for years the deep and passionate intellectual discussions Dr. Bertolasi generated to understand ischemic heart disease.

WHERE DID THE SCIENTIFIC PROOFS THAT SUPPORTED MEDICAL PRACTICE STEM FROM IN THAT PERIOD, THE BEGINNING OF THE ’80’S?

Cardiology had undergone a revolution during the sixties with the creation of coronary care units, cinecoronariography, coronary bypass surgery and echocardiography.

A new cardiological setting had emerged, with an overflow of new physiopathological concepts and the 24-hour hemodynamic and electrical evolution testimony of complex diseases in cardiovascular intensive care.

It was the time of physiopathology at the patient’s bedside, of small series experiments, and a shift of authority from experienced cardiologists to the new intensive care specialists who presented a previously invisible cardiology.

Bibliographic access was extremely difficulty: university libraries had been greatly damaged since the night of the long batons (during the military coup of 1966). Articles not available in Argentina required submitting the request to BIREME, established in Brazil, and patiently await the post until the precious article arrived three months later, an almost paleontological reference compared to the current Internet immediateness. Therefore, citing the New England Journal of Medicine or the American Journal of Cardiology (at that time the official organ of the American College, before the appearance of JACC) generated a halo of authority. Physicians of less complex centers were guided by the opinions of prominent figures belonging to more experienced centers. “The Argerich uses glucose-insulin-potassium, the Italian prescribes prazosin in heart failure, the Güemes places a Swan-Ganz in such or such patients”, were the greatest fountains of truth.

It was true that the coronary care unit had created a totally new experimental environment, where complex hemodynamic parameters as well as heart rhythm disorders could be repeatedly measured by means of accessible methods, allowing the quick assessment of different treatments. A good drug for heart failure was the one that decreased capillary pressure and increased cardiac output, and a good antiarrhythmic drug was the one that diminished the frequency of ventricular or supraventricular arrhythmias. Studies focused on a reduced number of patients, stressing the physiopathologicals effects.

WHERE DID THE SCIENTIFIC PROOFS THAT SUPPORTED MEDICAL PRACTICE STEM FROM IN THAT PERIOD, THE BEGINNING OF THE ’80’S?

Cardiology had undergone a revolution during the sixties with the creation of coronary care units, cinecoronariography, coronary bypass surgery and echocardiography.

A new cardiological setting had emerged, with an overflow of new physiopathological concepts and the 24-hour hemodynamic and electrical evolution testimony of complex diseases in cardiovascular intensive care.

It was the time of physiopathology at the patient’s bedside, of small series experiments, and a shift of authority from experienced cardiologists to the new intensive care specialists who presented a previously invisible cardiology.

Bibliographic access was extremely difficulty: university libraries had been greatly damaged since the night of the long batons (during the military coup of 1966). Articles not available in Argentina required submitting the request to BIREME, established in Brazil, and patiently await the post until the precious article arrived three months later, an almost paleontological reference compared to the current Internet immediateness. Therefore, citing the New England Journal of Medicine or the American Journal of Cardiology (at that time the official organ of the American College, before the appearance of JACC) generated a halo of authority. Physicians of less complex centers were guided by the opinions of prominent figures belonging to more experienced centers. “The Argerich uses glucose-insulin-potassium, the Italian prescribes prazosin in heart failure, the Güemes places a Swan-Ganz in such or such patients”, were the greatest fountains of truth.

It was true that the coronary care unit had created a totally new experimental environment, where complex hemodynamic parameters as well as heart rhythm disorders could be repeatedly measured by means of accessible methods, allowing the quick assessment of different treatments. A good drug for heart failure was the one that decreased capillary pressure and increased cardiac output, and a good antiarrhythmic drug was the one that diminished the frequency of ventricular or supraventricular arrhythmias. Studies focused on a reduced number of patients, stressing the physiopathologicals effects.

In summary, by the beginning of the eighties:
1. Physiopathological knowledge provided the rationale for the decisions.
2. The new cardiologists in the Coronary Care Units markedly influenced the trends of opinion, replacing the traditional cardiologist authority.
3. Much trust was placed in personal experience and small studies.
4. In debates, each contender chose the bibliography of the studies that supported his point of view, ignoring others.
5. Epidemiological observations were relied upon: “infarcts treated in this way have a better outcome”, with scarce understanding of a concept I will explain below.
6. A multivariate analysis was a technical feat: a 7-variable logistic regression took a whole night’s work together with a pile of sheets that summarized iterations.

A negative consequence of this view was the great heterogeneity of conducts for the same pathologies. A review published in the eighties, reported that in some countries, digitalis was prescribed for myocardial infarction in 60% of the cases, while in others it did not reach 5%, and similarly with other conducts.

A favorable effect was the high self-esteem, derived from believing that the physiopathology of diseases was understood, and the confidence in the relevance of personal investigations.

THE REVOLUTION OF THE 80’S

AND THE EXPLOSION OF THE 90’S

At the beginning of the 80’s few large cardiology studies had been performed; just the Coronary Drug Project on secondary prevention (1) with overall...
negative results, and initial cardiovascular surgery studies. (2) In 1984, what would become the megatrials manifesto and the origin of the later called “evidence-based medicine” was published. Salim Yusuf, Richard Peto and Rory Collins explained why we needed large studies to answer simple questions. (3) The Oxford group organized the ISIS I study on betablockers in infarction, (4) and then the ISIS II study, (5) with its counterpart in Italy, the GISSI I study, under the leadership of Gianni Tognoni. (6)

From different points of view, the proposal was conceptually new (Table 1).

Most interventions usually have a moderate beneficial effect. The authors showed that to unequivocally assess 20-30% reduction in myocardial infarction mortality, which at that moment was 12%, studies with more than 10000 patients were required. If their results were positive, they could readily be applied to medical conduct.

In addition, the conceptual role of meta-analysis was incorporated. The truth of researched information on a subject had to stem from the joint analysis of all the studies. With this hypothesis, applied to beta blockers and thrombolytics, they could predict that the sum of streptokinase studies suggested a 20-30% reduction in mortality that could be assessed in prospective trials.

When the prospective studies ISIS II in 16000 patients and GISSI I in 12000 patients with streptokinase were performed, they confirmed what the meta-analysis predicted: the drug notably reduced myocardial infarction mortality, which could not be proved with smaller trials. This started the myocardial reperfusion era.

Between 1987 and 1989, a series of meta-analyses were published reviewing each of the procedures used in acute and chronic ischemic cardiomyopathy: nitrates, beta blockers, lidocaine, aspirin, and heparin, and the joint view allowed for the first time the preparation of a sound systematic approach to treat myocardial infarction. (7, 8) Remarkably, the debate that seemed to go on forever regarding the utility of routine lidocaine in infarction to prevent ventricular fibrillation was buried after verifying in the meta-analysis, that it was associated with 33% mortality, in the limit of statistical significance. This sole observation led to the elimination of its routine use.

**THE EMERGENCE OF EVIDENCE-BASED MEDICINE**

David Sackett had been hired by the Rockefeller Foundation for the elaboration of concepts enabling a rational analysis of medical expenses and to judge medical practice validity through efficacy. (9) The convergence of this theoretical perspective with available information from large clinical trials and the widespread use of meta-analysis as an accessible tool for the community, gave rise to Evidence-Based Medicine (EBM).

In 1996, Sackett et al. published a definition in the BMJ (10) stating that EBM is the conscious, explicit and careful use of the best available evidence for decision making in the care of individual patients.

This new ideological definition created a new framework for clinical practice, providing a precise methodology to measure therapeutic effects as well as novel concepts (relative risk, odds ratio, relative risk reduction, etc., with their corresponding margins of error and confidence intervals). It became necessary, and thanks to the Internet feasible, to reassess each step of medical practice, exploring the scientific evidence supporting them.

The influence of large clinical trials and the EBM concepts modified the therapeutic scenario of the best studied pathologies and treatment heterogeneity between units or countries tended to disappear.

Fig.1 shows the comparison between discharge therapies in the CRUSADE Registry (11) in the United States and in an Argentine series generated by the Epi-Cardio project in 3555 myocardial infarctions. (12) There is a marked similarity in the indications, which is still currently used as quality control in intensive care.

**PRACTICAL LIMITS OF MEDICINE-BASED MEDICINE**

Creators of EBM describe an ideal scenario, in which a physician with sufficient technical knowledge poses daily questions on the clinical problems of his patients. Based on the framed question, he searches the best bibliography, processes the information extracting the quantitative data to express it in terms of EBM, and then makes the recommendation to the patient. Anyone who has tried to make that effort knows it demands long hours and that the personally obtained information must necessarily be compared with the intellectual work of other analysts on the subject to reach valid conclusions. The estimated weekly time a physician dedicates to reading is one hour. Therefore, in medical practice, the work of reviewing information lies in the hands of experts who transfer it in consensuses or guides, or publish it as evidence books.

It is true that each consensus establishes the levels of evidence and refers to all the literature, but in practice EBM has meant a new shift of the principle

| Table 1. Modification of clinical trials after the Oxford Group proposal |
|-----------------------------|-----------------------------|
|                             | Previous                | After/Posterior        |
| Nº of patients              | 10-100                   | >10.000                |
| Objective                   | Pathophysiological        | death                  |
| Centers                     | 1-2 high technology       | >150 communitary       |
| Efficacy criteria           | 40-50                    | 2 or 3                 |
| Design                      | Complex                  | simple                 |
| Clinical translation        | complex                  | simple                 |
| Researcher                  | Specialist               | physician              |
of authority to “evidentiologists” or clinical trial leaders. Numerous other limitations of this idea have been pointed out: the possibility of transforming these guides into mandatory demands, of their being adopted by healthcare systems with loss of medical autonomy, the impossibility of including in outcome measurements the true dimension of the doctor-patient relationship scenario, and many others.

We are going to discuss in some detail a criticism made from inside EBM on the conditions in which current investigations are generated, interpreted and released.

MODIFICATION OF RESEARCH SCENARIOS AND A NEW CRITICAL REALITY

The multicentric studies GISSI and ISIS had a small support from the industry, but were designed and carried out by independent groups and voluntary community networks. The success of these enterprises converted multicentric trials as the main methodology to assess new drugs for patenting, as well as the scientific basis for their application to the community.

Even though independent groups and public financed projects are still performed, most scientific production emerges from industry-sponsored trials. Table 2 summarizes some of the main changes since the 80’s up to the present.

Many clinical settings use drugs or perform interventions that have proved to reduce mortality. All present innovations cannot be compared with placebo but versus highly efficient drugs. The expectations of modifying mortality are statistically scarce and of small population impact, leading to the use of combined efficacy criteria (end points) for new therapeutics, as well as the non-inferiority design. Instead of mortality, current efficacy criteria are the combined event of cardiac death, non-fatal myocardial infarction, non-fatal stroke and in some cases absence of adverse events such as major bleeding.

In other cases, criteria as prevention of new diabetes with hypoglycemic drugs or of new hypertension with hypertensive drugs have been used, with scarce clinical utility criteria.

SOME RECENT EXAMPLES OF HIGHLY QUESTIONABLE USE OF INFORMATION OR OF DATA PRESENTATION

Subgroup manipulation

The CHARISMA study assessed the efficacy of clopidogrel and aspirin combination over aspirin alone in subjects with history of cardiovascular disease or in high risk groups. (13) The outcome was negative, i.e., no benefit was observed in the clopidogrel and aspirin combination group compared to aspirin alone, with the aggravating fact of a significant increase in bleeding. The authors formed a retrospective subgroup, unrealistically defined as “evident clinical atherothrombosis” which was compared to patients without that criterion. They detected a beneficial effect in the first subgroup whereas it resulted harmful in the second group, with increased vascular mortality. This type of regrouping is wholly invalid and criticized by all specialists in subgroup analyses, but the abstract published in a very prestigious journal concluded with two sentences:

The first one, which should have never been published, claimed:

In this trial, there was a suggestion of benefit with clopidogrel treatment in patients with symptomatic atherothrombosis and a suggestion of harm in patients with multiple risk factors.

The second sentence, expressed the real study outcome: Overall, clopidogrel plus aspirin was not significantly more effective than aspirin alone in reducing the rate of myocardial infarction, stroke, or death from cardiovascular causes.

The same journal devoted two highly critical editorials in relation to the abstract, suggesting that
allowing such a conclusion in the abstract was the result of negotiations with the editors. (14, 15)

**Questionable use of infarction definition**

The TRITON study comparatively assessed prasugrel with clopidogrel in patients with acute coronary syndrome referred for angioplasty. Plasugrel demonstrated a reduction of the combined end-point of death, myocardial infarction and stroke (16). The reduction was completely focused on the incidence of myocardial infarction most of which were periprocedural. (17) A review performed by the NICE consensus of Great Britain, pointed out that only 20% of the clinical signs of myocardial infarction were recognized by the physicians within the first few days. Expressed somewhat ironically, 80% of myocardial infarctions were not clinical, merely enzyme elevations. The Third Universal Definition of Myocardial Infarction published this year will only consider well-defined clinical signs and symptoms (pain, new electrocardiographic changes and/or persistent alterations in wall motion) as myocardial infarction after angioplasty; hence, most of the “enzymatic myocardial infarctions” of the TRITON study would disappear as events. (18)

An aggravating information on this study as well as on the PLATO study appeared in a recent publication by Serebruany et al. (19). The result is summarized in Figure 2.

In large studies it is fundamental to document the consistency with which the trial main events are diagnosed. Researchers declare events and provide documentation, as well as additional data which are sent to an event adjudication committee, which blindly ratifies or adds events to the study treatment. In the case of the TRITON study, the event adjudication committee doubled the number of myocardial infarctions compared to those detected by the researchers. Thus, the difference resulted statistically significant. In the case of the PLATO study, the event adjudication committee increased by 45 the number of myocardial infarctions in the group treated with clopidogrel and detected no under-diagnosed myocardial infarction with ticagrelor, turning a negative outcome into a statistically significant study.

Evidently, death and stroke with sequel are rarely readjudicated, but in these two cases the number of myocardial infarctions increased notably in comparison to those reported by the researchers, thus, positioning both studies in a significance level they had not previously reached. This is really serious and, under my impression, the result of manipulating the reported data and not of an improper conduct of the event adjudication committee integrated by impeccable professionals.

**Peculiar criteria to establish follow-up**

The results of the ATLAS study evaluating the efficacy of low dose rivaroxaban (2.5 mg every 12 hours) compared with placebo in follow-up after acute coronary syndrome treatment with aspirin plus clopidogrel were published in January 2012. (20) The study had a positive outcome with a significant small long-term event reduction. A very low rate of loss to follow-up, from 0.2 to 0.3%, was reported in the publication. However, when data were submitted for FDA drug approval assessment, the reviewers detected a loss to follow-up of 12% patients. (21) This corresponded to a novel criterion adopted by the trial: when patients definitely discontinued treatment they were only followed-up for a further month. Therefore, no long-term follow-up data of patients who discontinued treatment, a group that typically concentrates a significant number of events in clinical trials, was available. To discontinue follow-up implies to infringe the strict intention-to-treat criterion, a key tool to interpret pragmatically trial results. An absolute event difference of 1% to 2% with a 12% loss to follow-up is not reliable.

**MISCELLANEOUS WHICH OBSTRUCT TRIAL INTERPRETATION**

Some brief examples within the varied trial complexity of recent years:

**DREAM study. Tailored event and underestimation of a serious adverse event**

The benefit criterion was the prevention of new diabetes by assessing the use of a hypoglycemic drug (rosiglitazone) compared with placebo. The study reported a remarkable benefit, 60% reduction of new diabetes incidence. (22) Different reviewers pointed out that this new diabetes was simply a chemical finding, with a clinical relevance difficult to evaluate, and
they even pronounced the thoughtlessness of administering a drug preventively to avoid administering it later on, with no other clinical consequence. (23) The cardiovascular effect of the study drug was harmful. However, although the authors curiously claimed that cardiovascular events were similar in both groups, the event relative risk was 1.37 (95% CI 0.97-1.94) higher with rosiglitazone, at the limit of statistical significance. Further studies confirmed the cardiovascular risk, with very negative consequences for the continuity of the drug on the market.

OASIS study 7. Publication of a subgroup as a prospective trial.
The study evaluated the utility of a high dose of clopidogrel compared with a standard dose in patients with acute coronary syndrome. (24) The design implied admitting the patients to the clinic, where, according to the treating physician criterion, a group was referred for angioplasty and another continued with medical treatment. The study result was negative, with no advantage for high doses of clopidogrel. Retrospectively, the authors observed that in the subgroup of patients submitted to angioplasty the effect was beneficial, while it tended to be harmful in the conventional group. This is a valid analysis to generate a hypothesis, but contrary to expectations, the angioplasty subgroup results were published as a prospective design, concluding that the dose was effective. (25) When a study is negative, a positive effect on a subgroup cannot be claimed. It is merely a hypothesis to take into consideration. (26)

Meta-analyses with discordant results
In recent years we have read several meta-analyses on the same subject that yield different results, as in the case of cardiovascular risk with rosiglitazone. In general, independent analyses report variable results, whereas those financed by the industry are favorable to the drug. (27)

Clinically inaccurate definition of the treatment group
The SHIFT study assessed the efficacy of ibravudine in patients with betablockers or dose increase contraindications. The result was favorable, with a significant reduction in re-hospitalization or death. However, taking into consideration that only a few patients were treated with optimal doses, the question of whether these results would have persisted by adjusting the betablocker dose in patients already receiving them is still unanswered. (28)

THE CURRENT STATE OF EVIDENCES AND A COLLECTIVE ATTITUDE
The list of trial data and strategy manipulation may well be longer, and contributes to the complexity of arriving to conclusions in consensuses and guidelines. In recent years consensuses have increased the number of evidences based on experts’ agreement, i.e., not conclusive and dependent on the participants’ opinion, most of which have major conflicts of interest. (29)

Currently, there is a movement in institutions and medical journals to ensure greater transparency in clinical trial information. A previous protocol registry is demanded and some journals as for example JAMA and Annals of Internal Medicine even demand access to database to allow an independent review of the results. (30) In most cases, authors refuse this request and send their work to other journals.

An even more aggressive proposal on the future method of reporting clinical trial information has arisen from Richard Smith, former editor of the British Medical Journal. He suggests that the authors publish clinical trials financed by the industry in their own journals, that database is opened to the community, and that journals take the task of critically analyzing the conclusions through independent opinions. (31) This would keep great journals, historically prestigious, from supporting studies they have neither been able to examine nor evaluate in their central aspects.

To avoid biases and interests involved in the consensuses, the creation of interdisciplinary groups as for example Great Britain’s NICE, which includes physicians, specialists, general practitioners and nonmedical community participants, has been proposed. NICE’s recommendations are generally more conservative than those arising from the
consensuses of the scientific societies representing the specialty.

As specialists involved in the development of national guidelines for clinical practice, we have a great responsibility in this area. The first undeniable attitude is to maintain a prudent distrust of the published information. Ideally, before reaching definitive conclusions and communicating them to the community, not only published material but also the large amount of freely accessible information that is sent to the FDA for patenting, containing relevant aspects that are not published in ordinary journals should be analyzed.

The groups responsible for the local consensuses should not only evaluate this information rigorously but, even more importantly and from a broader point of view, estimate the relevance these data have for the practice of medicine in each community. In the nineties Gianni Tognoni envisioned an idea: properly organized medical practice might become the daily scenario for the assessment of therapeutic conducts. We are now a little closer to this possibility, taking into account the advances towards the possibility of a universal clinical history that allows assessing the merits and the real impact of interventions in the whole community. Although its realization is still a few years away, it is already technically feasible and will allow a new evaluation approach, today inaccessible, but never more necessary.

The nostalgic allusion to the revolutionary evidences “from the past”, talking about the eighties and nineties, only had the intention of contrasting it with the current complex situation, contributing to the debate of intricate and permanently reassessed ideas.

Many complementary steps must be accomplished to improve the strength of evidences: to focus the research on the patient and the community and not on the drug or the device, to consolidate a research that is independent and adjusted to medical systems and practices in the real world and to generate independent assessment structures associated to health authorities, among others. A no small contribution is to maintain independent criteria and to criticize the transgressions that are clearly unacceptable and harmful to patients and the community.

Dr. Carlos D. Taje
Director of the Argentine Journal of Cardiology

REFERENCES
1. The coronary drug project. Findings leading to further modifications of its protocol with respect to dextrothyroxine. The coronary drug project research group. JAMA 1972;220:996-1008.
2010;363:930-42.