JORGE THIERER

Catheter ablation in atrial fibrillation: similar or better than pharmacological treatment? The CABANA study and its different readings

Packer DL, Mark DB, Robb RA, Monahan KH, Bahnson TD, Poole JE, et al. Effect of Catheter Ablation vs Antiarrhythmic Drug Therapy on Mortality, Stroke, Bleeding, and Cardiac Arrest Among Patients With Atrial Fibrillation: The CABANA Randomized Clinical Trial. **JAMA 2019. 2019 Mar 15. doi: 10.1001/jama.2019.0693**

Mark DB, Anstrom KJ, Sheng S, Piccini JP, Baloch KN, Monahan KH, et al. Effect of Catheter Ablation vs Medical Therapy on Quality of Life Among Patients With Atrial Fibrillation: The CABANA Randomized Clinical Trial. **JAMA 2019 Mar 15. doi: 10.1001/jama.2019.0692.**

For years the treatment of atrial fibrillation (AF) consisted in the use of negative dromotropic drugs to achieve heart rate control or antiarrhythmic drugs (AD) for rhythm control, together with oral anticoagulation (OAC). In different randomized studies rhythm control demonstrated similar long-term results to frequency control, to which we must add the variable incidence of adverse effects from the use of AD, and a long-term efficacy that was far from optimal. In the last two decades we have witnessed the emergence and development of catheter ablation (CA) therapy for AF, as a rhythm control alternative that began to dispute AD treatment. Gradually, more complex patients were treated with this technique. However, beyond observational studies with evident selection bias (patients treated with CA when conditions were optimal) showing promising results, and randomized studies, also with very carefully chosen patients, suggesting better results with CA (less recurrence of AF, better evolution of the left ventricular ejection fraction), so far, there have been no large studies that confirm the advantage of CA over the usual pharmacological treatment. We have recently learned about the CASTLE AF study, which we discussed in RAC 2018; vol 86 nro 1. As we observed, this study included slightly over 300 patients from more than 3,000 initially considered, which reveals a very marked selection. Hence, the remarkable results of this study (the first to show mainly reduction of mortality) have not substantially modified the clinical practice.

Two publications of the CABANA study with a much larger number of patients have now appeared, comparing CA with the use of AD in patients with AF. Admission criteria were: two or more episodes of paroxysmal AF or at least one episode of persistent AF in the last 6 months (lasting ≥7 days or with need for electrical cardioversion after at least 48 hours of its initiation); and age ≥65 years or <65 years with at least one risk factor for stroke: hypertension, diabetes, previous stroke, heart failure or other cardiovascular problems. Patients were randomly assigned in a 1:1 ratio to CA (it was mandatory to perform isolation of the pulmonary veins and any additional technique was at the discretion of the treating physicians, who should have experience of at least 100 previous procedures), or to AD (with the recommendation to try initially frequency control and only if it failed, rhythm control). Patients who had previously been subjected to a failed CA, or those in which two AD treatments had failed were excluded from the study. Initially, the primary endpoint of the study was overall mortality. An expected 3-year mortality rate of 12% in the AD group and 30% decrease with CA was postulated. With these assumptions, considering a power of 90% and an alpha error of 5%, it would be necessary to include 3,000 patients. The secondary endpoint was a composite of death, disabling stroke, serious bleeding or cardiac arrest.

During the study, it was evident that the inclusion rate and the incidence of events were lower than expected. This led to the reversal of the endpoints: the secondary endpoint became primary, and follow-up was extended to 4 years. In this case, the number of necessary patients fell to 2,200. Overall mortality became the secondary endpoint, as well as the recurrence of AF, changes in quality of life and a composite of cardiovascular mortality and hospitalization. To evaluate the recurrence of AF a subgroup of patients received a monitoring device to record each symptomatic event, as well as to perform 24 and 96 hours recordings at prespecified intervals. A period of 3 months "clearance" was established after entering the study, during which a new procedure in the AC group and drug testing in the AD group could be performed in case of treatment failure, without considering these recurrences in the comparison. An intention-to-treat analysis was postulated as primary analysis. In addition, a per protocol analysis was also prospectively proposed (comparing the patients in the AC group who were actually subjected to the procedure with all the patients in the AD group), and a per real treatment study (comparing all the patients undergoing either treatment, regardless of whether they had been assigned by randomization or as a result from group crossover during the study).

A total of 2,204 patients from 126 centers in 10 countries were enrolled in the study (1,108 in the CA group and 1,096 in the AD group) between 2009 and 2016. Mean age was 68 years, 63% were men; slightly over 80% were hypertensive, and 25% were diabetic. Ten per cent had history of stroke or transient isch-

emic attack. Median CHA2DS2 Vasc score was 3, 43% of patients had history of paroxysmal AF, 47% had persistent AF and the rest permanent AF.

In the CA group, 90.6% of patients was actually subjected to the procedure in the month following randomization. In 215 patients (19.4%) a new ablation was necessary due to recurrence of AF during follow-up, and in 25 cases in the "clearance" period. Almost 45% of patients in this group received AD at some point during follow-up, and at the end of the study 26.5% of the patients in the CA group were medicated.

In the AD group, 99.6% actually received pharmacological treatment, in half of the cases with a single drug, and in 88.4% of cases, patients received treatment aimed at rhythm control. Throughout follow-up, 27.5% of the patients in this group underwent CA.

In the intention-to-treat analysis there was no difference in the incidence of the primary endpoint: 8% in the CA group, 9.2% in the AD group (HR 0.86, 95% CI 0.65-1.15, p=0.30). There was also no difference in total mortality between CA and AD (5.2% vs. 6.1%, respectively; p=0.38) but there was difference in the composite endpoint of death or cardiovascular hospitalization: 51.7% vs. 58.1%, respectively (HR 0.83, 95% CI 0.74-0.93, p=0.001). On the other hand, in the per treatment analysis, there was a significant risk reduction of CA compared with AD for the primary endpoint (HR 0.67, 95% CI 0.50-0.89) and for total mortality (HR 0.60, 95% CI 0.42-0.86). The same happened in the per protocol analysis at 12 months (HR 0.73 and 0.68, respectively, both significant). Regarding AF recurrence, the intention-to-treat analysis revealed that risk with CA was reduced to less than half of that observed with AD. If at the beginning of the study 57% of the patients had persistent or permanent AF, this value had decreased to 26% in the AD group and to 16% in the CA group at the end of follow-up.

In the CA group, the most frequent adverse events were minor hematomas (2.3%), pseudoaneurysms (1.1%) and cardiac tamponade (0.8%). In the AD group the most frequent events were thyroid disorders (1.6%) and proarrhythmia (0.8%).

Together with the publication of the main clinical results, a study on quality of life was reported. It was evaluated with two specific instruments, the AFEQT and MAFSI scores. The AFEQT score is obtained considering the answer to 18 of 21 specific questions on symptoms, daily activities, and treatment. The score ranges from 0 (total disability linked to the presence of AF) and 100 (total absence of disability). A variation in the score of 5 or more points implies a significant change in quality of life. The MAFSI score considers 10 items, and questions on the frequency and severity of the symptoms. For maximum scores of 40 for frequency and 30 for severity (the most severe symptoms in both cases) a decrease of 1.6 and 1.3 points, respectively, implied significant improvement. The scores reported were evaluated at baseline, at 3 and 12 months for the first year, and then every 12 months. On the other hand, a questionnaire that incorporates questions from these 2 scores and others traditionally used to assess quality of life (such as the SF 36 score) was administered at 6 months and then every 12 months. Considering the number of patients and a follow-up of to 60 months yielded a total of 20,461 questionnaires, of which 90% was effectively administered.

In the case of the AFEQT score, patients in the CA group had a baseline value of 62.9, and those in the AD group 63.1. After 1 year, the mean score rose in both groups to 86.4 and 80.9, respectively. The difference of 5.3 points between the two groups was statistically significant and implies a significant improvement in the quality of life with CA compared with AD. The improvement was more noticeable in the tertile of patients with worse baseline score: 7.7 points versus 5.3 in the middle tertile and 2.7 in the tertile with higher scores. At 5 years, the difference diminished (3.4 points) but remained significant. The improvement was also greater (in this case the decrease) for the MAFSI score in patients of the CA group, with a difference in the mean decrease of the frequency score of 1.7 points and 1.5 for the severity score, in both cases higher than necessary to understand that quality of life has improved.

The CABANA study illustrates some of the conflicting points regarding randomized studies. Although it is true that they eliminate the selection bias in favor of a certain conduct, it is also true that there is a previous selection, that of those who enter the study. Therefore, compared to patients in the real world, patients in clinical trials are younger and have a lower rate of comorbidities. Participating physicians are also selected based on their experience and they belong to centers that meet certain criteria for quality of care. In these studies it is therefore more possible that the tested strategies are effective, with a lower incidence of complications than in the real world. The difficulty in the inclusion of patients (just over 300 per year in more than 120 centers involves less than 3 patients per center per year) indicates how selected the included population is. The need to change the endpoint in the middle of the study and prolong the follow-up to reduce the number of included patients is another proof of what has been said.

The intention-to-treat analysis is the only one that preserves the purity of randomization, which precisely aims to distribute equally the known and unknown baseline characteristics of the patients, allowing the results to be attributed only to the intervention. In this analysis, CA was not superior to the use of AD. But it is also fair to recognize that patient crossover and the lack of compliance with what was assigned in the randomization make it difficult to interpret the results. In the CA group, 10% of patients did not undergo the procedure, almost half received AD at some point in the study and at the end of the study 1 out of 4 patients was medicated with one of these drugs. In the AD group, in turn, and mirroring the CA group, over 1 out of 4 pa-

tients crossed over to undergo CA, thus weakening in the intention-to-treat analysis any advantage that the procedure could have offered to the endpoint. In fact, the per protocol analysis or actual treatment shows the superiority of CA over the use of AD, but the advantage of randomization has already been lost: patients treated in one way or another are no longer similar to each other; for some reason they have not undergone the assigned procedure and have crossed over to the other group. When crossover from one group to the other is so marked (more than 25%) we find ourselves in a dilemma: how much of that reported in the intention-to-treat analysis represents reality? Perhaps understanding that strategies and not treatments are compared could reconcile both points of view: an initial strategy using AD is not inferior to that of implementing CA, but it is possible that due to treatment failure a significant proportion of patients has to fall into the other alternative achieving a better outcome. It is worth highlighting that the use of CA in the intention-to-treat analysis ensured less AF recurrence, and a lower incidence of cardiovascular hospitalization.

It should be noted that in the quality of life study, there was improvement in both groups, with advantage for CA. Due to the open nature of the study, in which each patient knows which group he was assigned to, it is very difficult to exclude a placebo effect, in which patients subjected to a more complex intervention can therefore feel that their symptoms have improved. But the notable difference in the recurrence of AF (to less than half with CA compared with AD) gives this finding a plausibility hint.

Are statins effective in those over 75 years? Results of a meta-analysis

Cholesterol Treatment Trialists C. Efficacy and safety of statin therapy in older people: a meta-analysis of individual participant data from 28 randomised controlled trials. **Lancet 2019;393:407-415**.

Different meta-analyses have confirmed that a reduction of 1 mmol/l (38.67 mg/dL) of LDL cholesterol by using statins translates into a decrease of approximately 20% in the risk of vascular events. This applies to men and women and across the spectrum of cardiovascular risk. However, doubts have been expressed on the efficacy of the treatment in people older than 75 years, and the futility of the intervention has been sustained beyond that age. We now know a new publication on the subject of CTT (Cholesterol Treatment Trialists') Collaboration, a group established in 1994, dedicated to the meta-analysis of statin randomized studies to define their safety and efficacy.

Twenty-eight studies with at least 1,000 patients included and a follow-up of ≥2 years were analyzed, 23 in which statins were compared with placebo or conventional treatment, while 5 compared two relatively more intense or less intense treatments with statins. A total of 186,804 patients were considered, divided for the analysis according to age in: ≤55 years (21%), 56-60 years (17%), 61-65 years (20%), 66-70 years (20 %), 71-75 years (15%) and the group of interest >75 years, which represented 8%. The group >75 years was more frequently represented in 4 studies carried out in patients with heart failure (20% vs. 4% of those <75 years) or on dialysis (3% vs. 2% of those <75 years). It should be recalled that in both clinical conditions statins have not shown to be effective.

Regarding the effect of statins (compared with control, or a more intense treatment compared with a less intense treatment) on the incidence of major cardiovascular events, the meta-analysis shows a reduction of 21% for each mmol/l of LDL cholesterol reduction (HR 0.79, 95% CI 0.77-0.81). There is a tendency for the effect to decrease with age (HR 0.75 in those ≤55 years, HR 0.87 in those >75 years), which is in the borderline of significance (p for trend=0.06). By eliminating the studies with heart failure and dialysis, where patients >75 years are overrepresented, this trend disappears, meaning that the beneficial effect of statins is not reduced with increasing age. When taking into account the presence or absence of established cardiovascular disease, it is verified that there is no difference according to age in the effect of statins in patients in secondary prevention, while in contrast the benefit clearly decreases in patients in primary prevention as they become older, turning non-significant in those >75 years.

When considering the effect on coronary events, the overall reduction for each mmol/l of LDL cholesterol decrease is 24%. Again there is a tendency to less effect as age increases, which is maintained even when the studies on heart failure and dialysis are eliminated; but in those ≥75 years, although smaller, the effect is still significant. Regarding the prevention of revascularization procedures and stroke (25% and 16% overall reduction for each mmol/L of LDL cholesterol, respectively) there are no differences according to age.

Finally, concerning the prevention of cardiovascular death and all-cause death (12% and 9% globally for each mmol/L of LDL cholesterol reduction, respectively), there is also a tendency to lose the effect with older age. But by eliminating heart failure and dialysis studies, which accounted for 53% of vascular deaths, this trend disappears. There is no effect of statins on both non-vascular or cancer deaths.

The discussion about the usefulness of statin therapy in the elderly is long-standing. The low inclusion rate in many studies, the coexistence of comorbidities that work against the ability of these drugs to improve the prognosis (such as end-stage renal failure and heart failure) are factors that may have contributed to the fact that the specific effect on patients over 75 years was not clear. In this sense, the meta-analysis we present is notoriously important, because in general terms it confirms the benefit of treating high-risk patients even when they are older. It should be noted,

however, that this clear effect is achieved in patients without the comorbidities described, and fundamentally in secondary prevention. The decrease of the benefit in the elderly may be due to the nature of the atherosclerotic phenomenon, and the growing influence that hypertension and kidney failure can have as age increases. Probably, this may be the reason that in the specific case of primary prevention the evidence is much less firm. In this sense, the STAREE study, currently underway, will undoubtedly contribute to give us an answer.

It should be taken into account, however, that although risk reduction is proportionately lower in the elderly than in young people, since the baseline risk of cardiovascular events is greater in the former, the number of events prevented per 1,000 patients treated is greater. Hence, although the controversy continues, it seems certainly appropriate to treat patients at risk regardless of age, taking into account in each case the expected benefit, tolerance and risk of adverse events.

Sedentarism, physical activity and their interaction when defining prognosis. An observational study in 150,000 subjects

Stamatakis E, Gale J, Bauman A, Ekelund U, Hamer M, Ding D. Sitting Time, Physical Activity, and Risk of Mortality in Adults. **J Am Coll Cardiol 2019;73:2062-2072**.

It is clear that physical activity carried out on a regular basis improves cardiovascular prognosis and that a sedentary behavior, usually defined as an energy expenditure of less than 1.5 METs (metabolic equivalent of oxygen consumption) in a sitting or reclining position for hours, entails an adverse prognosis. But it is true that we are not all the time active or sitting all the time (and if we are not sitting it may be because we are standing or doing physical activity, or, on the contrary, we are sleeping). There is published information that sedentary behavior is a predictor of poor prognosis when it exceeds 10 hours a day; and also that moderate to intense physical activity during >60 to 75 minutes a day nullifies the poor prognosis caused by a sedentary lifestyle. The study we present tried to evaluate the joint prognosis of physical activity and sedentarism, understanding that, when present, each one of them displaces the other throughout the 24 hours of a day.

This was a prospective cohort study that was carried out in New South Wales (Australia) in men and women ≥45 years. They were subjected to a structured questionnaire in which they were questioned about the daily time they were standing, sitting or sleeping, as well as the weekly time devoted to walking, and moderate or intense physical activity. The weekly physical activity was categorized as absent (0 minutes), insufficiently active (1-149 minutes), sufficiently active in the lower range (150-299 minutes), sufficiently active in the upper range (300-419 minutes), and recommended (≥420 minutes). The daily sleeping time was dichotomized for the analysis in ≤7 hours and >7 hours. Vital statistics of the patient's condition were used, and the data was adjusted for age, gender, diet, smoking, diabetes, self-reported body mass index, socioeconomic level, marital status, urban or rural residence and personal health assessment.

The survey analyzed the records of 149,077 participants and their relationship with evolution. Mean follow-up was 8.9 years for total mortality and 7.4 years for cardiovascular mortality. For each of the mentioned physical activity groups, total and cardiovascular mortality were evaluated according to the daily sitting time (<4, 4 to <6, 6 to <8 and ≥8 hours). There was a clear interaction between sitting time and time devoted to moderate to intense physical activity for all-cause mortality. The sedentary time was an independent predictor of total mortality in inactive or insufficiently active people. Among those with sufficient activity in the lower range, only a time of daily sedentarism ≥8 hours was associated with worse evolution. In those with greater physical activity, sedentary time was no longer a predictor of mortality. Regarding cardiovascular mortality, there was also more risk in those with greater sedentary time in the groups with less weekly activity, although there was no evidence of a clear dose response relationship.

When analyzing the effect on total mortality of replacing 1 hour per day of sitting for 1 hour of standing, a risk reduction of 3% was verified only among those who are sitting ≤6 hours a day, but not among those who are sitting longer. On the other hand, replacing 1 hour per day of sitting for 1 hour of walking or doing intense physical activity was associated with risk reduction, especially in those who are sitting >6 hours a day. Regarding cardiovascular death, each additional hour of sitting was associated with excess risk (7%) only among those who sit >6 hours per day; the replacement of a sitting hour for a standing hour was associated with a decrease in risk only for those who are sitting ≤6 hours a day; but the replacement of a sitting hour for one hour of moderate to intense activity was associated with risk reduction regardless of the amount of sitting hours.

The association of sedentarism with worse evolution has a physiopathological basis. Greater activation of inflammatory phenomena, obesity, decreased vasodilator capacity, endothelial dysfunction are some of the phenomena that explain it. Physical activity reverses the situation, and improves the prognosis. This analysis has the attraction of jointly considering the effect of sedentarism and that of physical activity on the vital prognosis. We usually read publications that take into account one or the other. But as the authors point out very well, there is a phenomenon of interaction between both of them: one variable influences in different ways the evolution according to the different strata of the other. If physical activity is enough, the adverse prognosis imposed by sedentarism is diluted,

until it disappears in those who are most active. If the time of sedentary time is shorter, it is enough to stand for a longer period to reduce the risk of events; if it is longer it will be necessary to do physical activity. It is true that those who are more sedentary may be less healthy, and that this may explains the worse prognosis; adjusting for the assessment of self-perceived health status at least partially corrects this issue. The message of this publication is clear: those who, due to their working conditions or personal inclination, spend much of the day sitting down should make an effort to compensate this condition with regular physical activity. As an additional observation, we would like to express our admiration for societies in which, as in this case, the evolution of almost 150,000 people who answer a questionnaire and are followed-up for almost 10 years can be described, with the possibility of adjusting the observations by their baseline characteristics. Will we ever be able to reproduce the experience?

The recalibration of the most used scores of cardiovascular risk improves their predictive capacity and equals their performance.

Pennells L, Kaptoge S, Wood A, Sweeting M, Zhao X, White I, et al. Equalization of four cardiovascular risk algorithms after systematic recalibration: individual-participant meta-analysis of 86 prospective studies. **Eur Heart J 2019;40:621-631**.

A score or clinical prediction rule is a tool that can increase the certainty in the diagnosis, prognosis or prediction of the response to a therapy taking into account in each individual the history, examination and complementary study data. Its origin lies, in general, in the determination of the best multivariate model that expresses the association of the predictor variables with the response. In the evaluation of a score two properties are considered: calibration and discrimination. The calibration of the model has to do with the ability to predict in each case a probability (according to the value of the covariates) that is a true reflection of the observed probability. If the estimated probability (number of observations over the total that will show the final point of interest)] coincides with the one observed, the score calibrates adequately.

Discrimination refers to the ability to assign to the individuals who effectively present the condition or event of interest a greater probability of its occurrence compared to those without that condition, and it is evaluated through the C index, which corresponds to the area under the ROC curve. A prognostic model that adequately evaluates the risk of events helps to make therapeutic decisions. In fact, the indication of statins and the choice of a high or moderate intensity treatment are based on the use of different models.

Until 2014, more than 100 cardiovascular risk models had been published. Less than a third of them had external validation. And when there was such validation, the performance was not similar to that of the original model. Some examples are well known: the Framingham score, to mention one of the most marked cases, overestimates the risk of cardiovascular events in the British or Spanish population (poor calibration). It is true that in that case re-calibrating the score in the new population (with its own characteristics) can lead to a modification of the test to return its usefulness (for example, the REGICOR score in Spain). Different scores strongly recommended by the practice guidelines may differ subtly in the primary event they predict, in the predictor variables or in the way they have been modeled. The comparison between different scores applied to the same population to assess the primary indication of one or the other is very rarely carried out. That is why the publication we present is valuable.

Four commonly used scores were considered to assess cardiovascular risk: the Grouped Cohort Equations, recommended by the American College of Cardiology and the American Heart Association; the SCORE Risk Charts, recommended by the European Society of Cardiology; the Reynolds score and the traditional Framingham score. The 4 models were applied to 360,737 participants without previous cardiovascular disease, from 86 prospective cohort studies carried out between 1963 and 2003, and the ability to calibrate and discriminate was defined. None of these cohorts had been used for the preparation of the evaluated scores. The primary endpoint was the 10-year risk of cardiovascular events specifically considered for each score: acute non-fatal myocardial infarction (AMI), fatal coronary event or any stroke in the case of the Framingham score and the ACC/AHA score; the same plus coronary revascularization or any cardiovascular death for the Reynolds score and cardiovascular death for the SCORE Risk Charts. As a secondary endpoint for all scores, a cardiovascular event common to all four was considered: the aforementioned primary endpoint for the Framingham and ACC/AHA scores. Each score was recalibrated by modifying the prognostic algorithm to take into account the profile of risk factors and the observed incidence of events in each cohort, thus adjusting the predicted risk to the observed one. On the other hand, to allow head-to-head comparison between the four scores, the SCORE Risk Charts and the Reynolds score were recalibrated to predict the common cardiovascular event.

Mean age of participants was 59 years; 53% were men; 69% were European, 18% North American, and the rest from Japan and Australia. Median follow-up was 10.2 years. The median cardiovascular risk estimated at 10 years was 5.5% with the Framingham score, 2.5% with SCORE Risk Charts and 6.4% with the ACC/AHA score. The ability to discriminate of the different scores and for the various endpoints considered did not differ greatly, with a C index that ranged between 0.70 and 0.76; nevertheless, there was some superiority of the ACC/AHA and SCORE Risk Charts scores (areas under the ROC curve between 0.003 and

0.013 higher) over the Framingham score. None of the 4 scores adequately calibrated when estimating the risk of the event specifically predicted by each model with what actually happened: globally, considering all the cohorts, the Framingham score overestimated risk by 10%, the ACC/AHA by 41%, the SCORE Risk Charts by 52%, while the Reynolds score underestimated it by 10%. The SCORE Risk Charts and ACC/AHA scores made the same calibration error in men and women and throughout the age range; the Framingham score overestimated risk in young men and women, and underestimated it in elderly women. When considering the 86 cohorts separately, there were cases of risk overestimation of up to 400% and of risk underestimation of up to 50%.

However, after the recalibration of the 4 scores, adjusting the prognostic algorithms to the distribution of risk factors in each cohort and targeting the proposed common endpoint (non-fatal AMI, fatal coronary event or any stroke), the performance of the 4 scores was similar. Prior to the recalibration, 32% of participants evaluated with the Framingham score, 29% with the SCORE Risk Charts, 39% with the ACC/AHA score, and 32% with the Reynolds score had been considered at high risk, and therefore with an indication to start statins. After the recalibration the respective figures fell to 22%, 22%, 24% and 23%, respectively. There were no significant differences among scores in the ability to discriminate.

The construction of a score has 2 basic stages: the derivation from the original cohort, and the validation in other cohorts. Most of the time the authors are satisfied with publishing the score built from the derivation cohort, and do not validate it. This happens, as we said, with approximately 2 out of 3 prognostic models that are exposed to the consideration of the scientific community. When the validation is carried out, it is often performed in another cohort which comes from the original score population, so that the results are not very different from the original ones. But a score that seeks to be of usual clinical use should have been validated in other populations. There arises the problem of bad calibration. This is explained by the fact that the model is used in a population with different baseline conditions (in terms of the distribution of risk factors, socioeconomic conditions, access to treatment, etc.) and different incidence of events. The 4 scores considered in this publication have been widely validated; note, however, how, applied to the same populations, they differ in their predictive capacity. This happens because their algorithms are different, the prognostic variables and the predicted endpoint are not exactly the same, and their origin is not similar. When the recalibration is carried out taking into account the characteristics of the population to which they are applied, and they are modified so that they predict exactly the same endpoint, the differences disappear; and the proportion of high-risk patients is lower, since they generally overestimate the risk.

The findings are interesting from a methodological point of view, but at the same time they illustrate the weaknesses of the prognostic models: the changing conditions of the populations to which they are applied make predictive certainty illusory when the prediction is based on conditions that are no longer the same. The task of recalibrating scores is difficult to carry out periodically and involves significant use of resources, but it is an attempt to achieve greater closeness to the truth. Only knows reality he who is capable of predicting it ... rightly.

Level of evidence of the recommendations of practice guidelines. Has anything changed in the last 10 years?

Fanaroff AC, Califf RM, Windecker S, Smith SC, Jr., Lopes RD. Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. **JAMA 2019;321:1069-1080**

Clinical practice guidelines (CPG) have become an unavoidable part of the diagnostic and therapeutic process of contemporary medicine. There is enough evidence that following them translates into better results in patient care. We know that each recommendation has two components: the strength of the recommendation and the level of evidence to support it.

Regarding the strength of the recommendation, it can be I, II or III. For I it is understood that there is evidence and/or general agreement that a certain treatment or procedure is beneficial, useful and effective and we will say that the recommendation must be carried out. For II there is conflicting evidence and/or divergence of opinion about the usefulness or effectiveness of the treatment or procedure; it can be IIa, if the weight of the evidence or opinion is in favor of the procedure or treatment or IIb if the usefulness is less established, so that IIa should be considered and IIb, might be considered. For III there is evidence and/or general agreement that a certain treatment or procedure is not useful or effective, and in some cases it can be harmful. That is why it should not be carried out.

Regarding the level of evidence, it can be A (the evidence comes from 2 or more randomized studies, or meta-analysis), B (the evidence comes from only 1 randomized study or from large non-randomized studies), or C (the recommendation arises from consensus of experts' opinion, and/or from evidence of small, retrospective studies or registries).

In 2009 a well-commented publication noted that in the CPG issued by the AHA/ACC only 11% of the recommendations were based on level of evidence A. There were many requests then to improve that situation. What has happened since then? A group of researchers considered all the cardiac CPG issued by the AHA/ACC and the European Society of Cardiology between 2008 and 2018 and tabulated the recommendations according to strength and quality of evidence.

In the aforementioned period, 26 CPG from the AHA/ACC were published on different topics, with a

total of 2,930 recommendations: 43.4% were class I, 45.7% class II and 10.9% class III. In turn, 8.5% were based on level of evidence A, 50% on evidence B and 41.5% on evidence C. Among the firmer recommendations, I and III, only 12.9% were based on level of evidence A, 48.9% on evidence B and 38.2% on evidence C.

In the same period, 25 CPG of the European Society of Cardiology were published, with a total of 3,399 recommendations: 47.7% were class I, 44.6% class II and 7.7% class III. In this case, 14.2% were based on level of evidence A, 31% on evidence B, and 54.8% on evidence C. Among the strongest recommendations, I and III, 21.3%; were based on level of evidence A, 29.1% on evidence B and 49.6% on evidence C.

Among 16 CPG issued by the AHA/ACC, a guideline published before 2008 referring to the same topic was found. The median of recommendations based on level of evidence A did not vary when comparing the 2008-2018 period with respect to the previous one: 9% vs. 11.7% On the other hand, there was an increase in the recommendations based on level of evidence B (51% vs. 41.9%), and therefore a decrease of those based on evidence C (36.7% vs. 51.9%).

In the case of 16 European guidelines published between 2014 and 2018, a guideline on the same topic published between 2004 and 2014 was found. In this case there was no significant difference between the periods considered in the proportion of level of evidence A, B or C.

When it comes to following guideline recommendations, we expect that the most categorical (what must and must not be recommended) should be based on the strongest and least controversial evidence, of the highest quality according to Evidence-Based Medicine guidelines. It is therefore interesting that in the American guidelines just over half of the recommendations are I or III, but only 13% are based on level of evidence A (and even if this were not available and evidence B were looked for, it then turns out that almost 40% relies on expert consensus or small studies or registries). And in the European guidelines the situation is hardly better, with a similar proportion of recommendations I or III, with more level of evidence A (21%) but also more evidence C (almost 50%).

How can we read this data? In principle, it seems clear that there is not so much evidence "of maximum purity" when it comes to making recommendations. In fact, and going to the most elementary actions; do we know of any randomized trials on the need to take blood pressure or perform an ECG or an echocardiogram? The vast majority of everyday actions find their support in the empirical, not in the clinical trial. And would someone discuss the inexcusable indication of carrying out the aforementioned actions? It seems therefore irremediable that there are and will continue to be many class I recommendations with level of evidence C. However, some actions that at some time were certainly class I or III have fallen into disrepute, sometimes due to observation, sometimes to some clinical trial that dared to challenge the norm. Where is the indication for prophylactic lidocaine in infarction, the contraindication for beta-blockers in heart failure? This means that in many cases the level of evidence C could be removed and replaced by higher evidence if there were more frequent challenges and doubts. In conclusion, it could be established that there is little evidence A in the guidelines ... because there is indeed little evidence A. Randomized studies are mostly driven by the industry, and therefore generally refer to the use of drugs and devices. Studies that try to answer big questions about clinical management strategies beyond a particular therapeutic agent are scarce goods. But, to conclude, a difficult question to answer: why do American guidelines find evidence for their recommendations in 8% of the cases, and the European ones in almost double, 14%? Do they consider different sources, differ in the clinical trials assessed or value the same evidence in a different way? Even what seems indisputable (a certain level of evidence is A, or B, or C) may not be so. Perhaps, and under their appearance of biblical command, the recommendations of the guidelines also lie on interpretation.

Association between sugar-sweetened and artificially-sweetened beverage consumption and mortality in two large cohort studies

Malik VS, Li Y, Pan A et al. Long-Term Consumption of Sugar-Sweetened and Artificially Sweetened Beverages and Risk of Mortality in US Adults. **Circulation 2019;139:2113-2125.**

The consumption of sugar-sweetened beverages (SSB) is the most important source of sugar in the diet. It represents 6% of the overall caloric intake in adults in the United States. In epidemiological studies, this consumption has been linked to an increased risk of obesity, type 2 diabetes, coronary heart disease and stroke. The relationship with all-cause mortality or with cardiovascular mortality is less clear, with some studies that support the existence of this association and others that question it. In the case of artificially-sweetened beverages (ASB) there is less information. To clarify this point, an analysis of two large cohorts that have provided relevant information over the last decades was carried out: the Nurses Study, which began in 1980, and included 121,700 women between 30 and 55 years, and the Health Professionals Study, which began in 1986 and included 51,529 men between 40 and 75 years. In both studies, among other baseline determinations, a questionnaire was administered defining the characteristics of the daily diet in each participant, including questions specifically directed to the consumption of SSB and ASB. Participants who presented with diabetes, cardiovascular disease or cancer at the time of inclusion were excluded from this analysis, leaving 80,647 women and 37,716 men. For each beverage the following categories of consumption were established: <1 per month, 1-4 per month, 2 to 6 per week, 1 per day or 2 or more per day.

Throughout follow-up (34 years among women, 28 years among men) there was progressive decrease in the consumption of SSB, and an increase followed by a decrease in ASB consumption. Those with higher SSB consumption were younger, less physically active and smoked more. They also consumed more red meat and had greater glycemic load, and consumed less whole grains and vegetables. Those with higher ASB consumption were also younger, with higher body mass index and lower glycemic load. The increasing consumption of SSB, adjusted for age and ASB consumption, was associated with all-cause mortality. Compared with the lowest category, the highest category presented a HR for overall mortality of 1.52 (95% CI 1.43-1.61). The risk was greater in women than in men (HR of 1.63 and 1.29, respectively). After adjusting for family history, ethnicity, coronary risk factors, diet and body mass index, the relationship was attenuated but remained significant (HR 1.21, 95% CI 1.13-1.28). The excess risk associated with the increase of an additional serving (can, glass or bottle according to the drink) was 7%, and after adjusting for the abovementioned factors, this was 5%. Interaction with sex was verified, with a greater influence in women than in men. Increasing the consumption of SSB was also associated with higher cardiovascular mortality (adjusted HR 1.31, 95% CI 1.15-1.50) and an increased risk of death from cancer (adjusted HR 1.16, 95% CI 1.04 -1.29), especially risk of death from breast cancer.

Artificially-sweetened beverage consumption showed a lower association with overall and cardiovascular mortality in women: HR of 1.1 and 1, respectively for the highest category of consumption in relation to the lowest, which were attenuated after multivariate adjustment. When a category of at least 4 times per day was considered, the relationship became stronger, with an excess risk of 30% for overall mortality and 43% for cardiovascular mortality. There was no association with overall or cardiovascular mortality in men. The analysis allowed us to estimate that replacing a serving of SSB with that of ASB would result in a reduction of 4% in overall mortality, 5% in cardiovascular mortality and 4% in cancer mortality.

The association between SSB consumption and overall mortality, evidenced in this analysis of 2 large cohort studies, confirms previous findings and is biologically plausible, given the adverse metabolic effects that high carbohydrate load can produce, with a higher risk of obesity, diabetes, activation of inflammatory phenomena and vascular disease, in addition to the relationship shown with certain types of cancer. The association of ASB consumption with mortality is less convincing. In principle because there is still no clear pathophysiological explanation. There are those who sustain that the sweet taste of these drinks would stimulate the consumption of foods that are also sweet but not dietetic However, there is no evidence that consuming ASB leads per se to weight gain. The reverse causality (ASB are consumed by the hypertensive and obese, therefore, with increased insulin resistance, and eventually with a higher risk of diabetes) cannot be completely overruled. On the other hand it is striking that this relationship has been shown only in women, and only with high consumption, so the association looks less firm, and should be confirmed in new studies.

Similarities and differences of the effects of SGLT2 inhibitors and GLP-1 agonists in the evolution of diabetic patients. A meta-analysis of randomized studies

Zelniker TA, Wiviott SD, Raz I et al. Comparison of the Effects of Glucagon-Like Peptide Receptor Agonists and Sodium-Glucose Cotransporter 2 Inhibitors for Prevention of Major Adverse Cardiovascular and Renal Outcomes in Type 2 Diabetes Mellitus. **Circulation 2019;139:2022-2031.**

In the last 4 years we have witnessed a revolution in the field of diabetes: after decades in which the success of hypoglycemic treatment seemed focused on the reduction of blood glucose and microvascular events, we now know drugs (SGLT2 inhibitors or gliflozins and GLP-1 agonists) capable of modifying the vital prognosis and specifically decreasing the incidence of major adverse cardiovascular events (MACE): cardiovascular mortality, non-fatal myocardial infarction (AMI) and non-fatal stroke. Even empagliflozin and liraglutide have been shown to reduce all-cause mortality. Different mechanisms have been proposed by which these drugs have a benefit: natriuresis and renal protection in the case of glifozins (a fundamentally hemodynamic effect) and a predominantly metabolic and anti-atherosclerotic effect in the case of GLP-1 agonists. Are the patients´ prognoses similar as a result of using one or another type of drug? A meta-analysis of randomized studies answers this question. It considered 8 randomized studies. In 5 of them, GLP-1 agonists were tested (ELIXA with lixisenatide, EXSCEL with exenatide, SUSTAIN 6 with semaglutide, HARMONY with albiglutide and LEADER with liraglutide), and in 3, gliflozins (EMPA-REG Outcome with empagliflozin, CANVAS with canagliflozin and DECLARE with dapagliflozin).

A total of 77,242 patients were included, 55.6% of them in studies with GLP-1 agonists. Mean age ranged between 60 and 65 years and the proportion of women was between 28% and 40%. The prevalence of established atherosclerotic disease was 73.1%, but ranged between 41% and 100% according to the inclusion criteria and that of heart failure was 16.3%, (between 10% and 24% according to the studies). The prevalence of kidney failure defined as glomerular filtration rate <60 ml/min/1.73 m2 varied between 20% and 29% except in the DECLARE study, where it was only 7.4%.

The effect of both types of drugs on the incidence of MACE was similar: HR 0.88, 95% CI 0.84-0.94 for GLP-1 agonists and HR 0.89, 95% CI 0.83-0.96 for gliflozins. In fact, the effect was concentrated in patients

with established vascular disease, in whom the reduction reached 14% and was similar with both drug families, while in those patients with only the presence of risk factors the effect of both types of agents was non-existent.

Both GLP-1 agonists and glifozins significantly reduced the risk of AMI: 9% GLP-1 agonists, 11% gliflozins, without evidence of heterogeneity between the two drugs. Similarly, both drugs reduced the risk of cardiovascular death: 12% GLP-1 agonists, 16% gliflozins. Only GLP-1 agonists decreased the risk of stroke by 14%, while there was no significant reduction with gliflozins. On the other hand, only the latter significantly decreased by 31% the risk of hospitalization due to heart failure. Although both families of drugs decreased the incidence of a composite endpoint of renal events, the effect was more marked with glifozins, with a reduction of 38% compared with 18% with GLP-1 agonists. In addition, there was a difference in the specific effect: GLP-1 agonists focused on a reduction in the incidence of macroalbuminuria and glifozins in the duplication of creatinine levels, the incidence of end-stage kidney failure and renal death.

This meta-analysis of large randomized studies confirms what is already known about the favorable effect of both types of drugs on MACE. It confirms that it focuses on patients who already have established cardiovascular disease, and it is not verified in those who only have risk factors. However, the REWIND study with dulaglutide, not yet published, showed a reduction of MACE in patients with risk factors free of proven cardiovascular disease, but at a follow-up of 8 years, suggesting that it may take more time to achieve significant effects in earlier stages. The beneficial effect of glifozins on the incidence of heart failure and kidney disease has already been adequately evidenced, as well as the fact that the reduction of stroke is expected with GLP-1 agonists and not with glifozins. But what draws our attention is that glifozins and GLP-1 agonists similarly reduce the incidence of AMI. If until now it was postulated that the anti-atherosclerotic effect was specific of GLP-1 agonists, and that the effects of glifozins were fundamentally hemodynamic, how to explain this point? It was expected that the reduction of the rate of AMI would be an aspect in which the GLP-1 agonists would take advantage. But it should not be forgotten that, due to its glucosuric effect, SGLT2 inhibitors also exert an anti-inflammatory action and, therefore, anti-atherosclerotic action, which could be involved in the described findings, and that the effects of both families of drugs are so varied that a univocal explanation is undoubtedly illusive. It will not be the first time that the clinic helps to interpret the physiopathology.