EDITORIAL

The ARUBA study: where do we go from here?

Mohamed Samy Elhammady, MD, and Roberto C. Heros, MD

Department of Neurosurgery, University of Miami, Florida

The highly controversial ARUBA (A Randomized Trial of Unruptured Brain Arteriovenous Malformations) study marked an important turning point in the history of the management of unruptured cerebral arteriovenous malformations (AVMs)\(^4\). This issue of the Journal of Neurosurgery includes a very thoughtful and important paper by Magro and colleagues, addressing the various critiques that have been published since the inception of the study in 2006.\(^3\) However, before discussing the current paper, we would like to briefly summarize the design and results of the ARUBA trial.

The ARUBA study was the first prospective multicenter randomized controlled trial (RCT) comparing medical management alone to medical management and prophylactic interventional therapy (endovascular, surgical, and radiation therapy, alone or in combination) in patients with unruptured AVMs. The primary end point was the composite measure of death from any cause or any stroke, which was defined as any symptom associated with any imaging finding. The secondary analysis was overall functional status and quality of life at a minimum of 5 years from randomization. An interim analysis performed after 6 years led to discontinuation of the study due to excessive morbidity in the treatment arm as compared with the conservatively treated cohort. A total of 223 patients had been enrolled in the trial, with a mean follow-up of approximately 33.3 months. The primary outcome was seen in 11 (10.1%) of the 109 patients assigned to the medical group and in 35 patients (30.7%) of the 114 patients assigned to the intervention group.

Even prior to commencement of enrollment, the design of the ARUBA study had been heavily criticized, particularly in regard to the proposed 5-year follow-up period, which many argued would unfairly detect all procedure-related complications but would be too short to detect the potential long-term benefits of prophylactic intervention. It was therefore not a surprise to most when the study was prematurely interrupted and the results were in favor of conservative therapy. Numerous editorials and commentaries discussing the weaknesses of the trial soon followed.\(^1\) However, despite the widely discussed limitations, many have interpreted the results as indicative that no unruptured AVM should be treated at all.\(^2\)

In this issue of the Journal of Neurosurgery, Magro and colleagues performed a systematic review of all articles and comments published between January 2006 and February 2015 that discussed the ARUBA study. After excluding articles unrelated to ARUBA and those written by the ARUBA Collaborative Group as well as those not containing detailed comments on the trial, a total of 31 articles were selected for analysis. The authors then categorized the critiques into 3 broad categories: those related to study design, conduct, and analysis and interpretation. In regard to study design, the heterogeneity of the patients and the lack of standardization of the treatment arm were mentioned most frequently, followed by the choice of outcome measures and the primary hypothesis in favor of medical management. In regard to conduct of the study, the low enrollment rate and the suspected selection bias were the most frequently mentioned concerns, as well as, to a lesser extent, the premature interruption of enrollment. Regarding analysis and interpretation of the trial, the short follow-up period and inappropriately drawn conclusions were the most frequently cited critiques, followed by the lack of subgroup analyses and the lack of detail regarding the treatment results. The authors discussed in a very thoughtful manner each of these limitations and subsequently proposed a framework for future clinical trials that would avoid some of the shortcomings of the ARUBA study. We would like to take this opportunity to highlight some aspects of their discussion.

The authors point out that one of the main problems
with the ARUBA study had to do with the trial hypothesis: that conservative management improves patient outcomes as compared to prophylactic intervention. They explain how this hypothesis is not only unusual but illogical because intervention should have been the experimental arm that needed to be tested against conservative management, which should have been the control group, and not vice versa. This is because patients undergoing any sort of intervention are taking the risk of the procedure upfront, and in the setting of a still unproven treatment it becomes incumbent upon intervention to demonstrate the long-term benefit. Similarly, in the absence of evidence that an intervention is more beneficial than medical therapy, there is no need to test conservative management. In fact, it is unethical to presume that intervention is harmful and inferior to medical therapy and then decide to prove it in a trial. Another issue that the authors stress is that interventions must be tested individually to prove a beneficial effect based on their own merits. The results obtained from testing a group of interventions without stratification cannot be used as validation for or against any of the interventions individually. This is because a beneficial intervention may not be able to prove itself when lumped together with a harmful intervention and vice versa. Thus in the ARUBA trial, where all types of intervention were grouped together, only conservative management could be tested and verified. Based on these 2 points, the authors conclude that the ARUBA study ended up testing the single option that in the absence of evidence did not need to be tested.

Magro and colleagues argue that the only reason why the ARUBA study was able to take place was due to failure of our profession to prove the beneficial effects of the interventions we adopt. There are undoubtedly many changes taking place in our health care system. Given the rising cost of medical care, there has been a strong demand from government and health insurance agencies for evidence-based medicine and outcome measures. The burden of proof has fallen on our shoulders to prove the effectiveness of our interventions. The authors make the point that we, as a medical community, must learn how to better integrate clinical research and medical care and that the ARUBA trial should serve as the starting point for a collaborative effort by our profession to practice medicine in a more scientific fashion through so-called care trials. If we were to presume that conservative management is the current standard of care for unruptured cerebral AVMs, the authors propose establishing a new care trial whereby a promising yet still unproven intervention is offered to patients and tested for its merits against medical therapy (the standard of care). It is important to note that no patient should be treated outside of the scope of a care trial. If patients refuse possible randomization to intervention, then they must receive conservative management. The proposed study can still remain pragmatic and practical so as to minimally affect clinical care, but should consist of stratified treatments and involve a large number of patients. Similarly, the primary outcome should be well defined (e.g., mRS score > 2) and measured at a time interval that is sufficient (5–10 years) to demonstrate the potential benefits of therapy.

Unfortunately, and as the authors themselves acknowledge, such long-term care trials involving a relatively uncommon pathological entity can be a difficult undertaking for reasons such as inability to secure long-term funding, an increased vulnerability to post-randomization confounding and selection biases, and a greater potential for losses to follow-up and patient cross-overs. Another problem, which the authors do not address, is the issue of clinical equipoise. Is it ethical, at least in the mind of a competent cerebrovascular surgeon, to randomize a young patient with a small anterior frontal pole AVM to conservative management and thus expose him/her to the definite, albeit small, risk of hemorrhage, and deny him/her the chance of a cure through resection, which can be performed with minimal morbidity? Conversely, would we randomize an older patient with a Spetzler-Martin (S-M) Grade V AVM knowing that any intervention in this setting is ineffective, very dangerous, or both? Do we disregard personal experience and the wealth of literature accumulated over the years because it is was not obtained through RCTs? Finally, we must keep in mind that cerebral AVMs are uncommon complex lesions and that no two patients are similar. There are many confounding factors that influence both the natural history and the risk of intervention, and thus it becomes impossible and impractical to conduct a study that can account for every possible combination of variables.

If we accept that it would be difficult and impractical to conduct a proper RCT as suggested by Magro and colleagues, is there an intermediate, practical course of action that may improve the overall care of patients with unruptured AVMs? We believe there is. We can assemble a panel of experts who can embark on a systematic and thorough review of the extensive published experience available, much as we have done to establish guidelines in other areas of neurosurgical care. We are convinced that, even in the absence of proper randomized studies, there are enough published data concerning both the results of treatment with each of the different modalities on the one hand and the natural history of untreated AVMs on the other, that this process is likely to result in a general classification of AVMs in 3 broad groups. One group, which is likely to comprise most S-M Grade I and II AVMs, would probably be found to generally benefit from treatment. Another group, including most S-M Grade IV–V AVMs, will probably be best left untreated given the available data. And certainly we will find that in a significant number of patients, probably including most of those with S-M Grade III AVMs, some of the most difficult Grade II AVMs, and the easiest Grade IV AVMs, we simply do not know from the data available whether intervention should be undertaken, and what form of intervention (if any) should be offered. This will be the group in which true clinical equipoise exists, and the group of patients that most experienced clinicians would be likely to recruit into a well-designed randomized trial that adheres to the requirements so well discussed by Magro and colleagues.

We are indebted to these authors for what we consider to be the most thoughtful critique to this date of the ARUBA study.

http://thejns.org/doi/abs/10.3171/2015.7.JNS151408
Furthermore, comparing surgical results (in patients we do want to treat) with the “natural history” (of the patients we do not want to treat) is undoubtedly an invalid procedure, which has been responsible for major errors in the past.\textsuperscript{2,7,11} Consider the era prior to the extracranial-intracranial (EC-IC) bypass trial: “There is certainly a need for more investigations about the natural history of the disease and the comparison with the long term follow-up results in surgically treated and untreated patients. However, in our experience, the natural fate of patients is unhappy and the treatment seems capable of modifying the natural course of the disease,”\textsuperscript{13} and, “[Surgery] is a reasonably safe and technically satisfactory procedure that has a potential of improving outcome, as compared with the natural history of the disease.”\textsuperscript{12} The preceding “expert opinion” quotes are 30 years old; these were revised after the publication of the EC-IC bypass trial in 1985.\textsuperscript{5} Medical history has shown time and again that expert opinions and biased data from observational studies can be overturned when RCTs are finally performed.

That said, the systematic review of observational studies of treatment options is an excellent first step to prepare hypotheses, but hypotheses that must be rigorously tested within trials assessing the real world of patient care.

Let’s take Elhammady and Heros’ proposal one step further and imagine that a mega-trial on all brain AVMs has now been completed. In this imaginary trial, patients were randomly allocated to conservative management or surgery groups; the primary outcome was the proportion of independent patients (mRS score < 3) at 10 years. The results are illustrated with a forest plot (Fig. 1). This trial result would justify the “3 broad groups” that Elhammady and Heros propose. Patients with small or low-grade AVMs are shown to do better with surgery; S-M Grade IV–V AVMs do better with observation; results for mid-size or S-M Grade III AVMs remain unclear.

We would like to emphasize 2 points. First, the results of our imaginary trial are very improbable. It is extremely unlikely to obtain clear-cut, significant differences in treatment outcomes, in diverging directions according to patient and AVM type (note the arbitrary age and size cutoffs with diverging results). If this may make sense for a minority of patients at the extreme ends of the various spectra (i.e., S-M Grade V AVMs that cannot be resected), for experts to make such treatment decisions based on patient characteristics is equivalent to assuming that such a trial has been conducted and such improbable results are in fact reality.

The second point concerns the patients in the intermediate “gray zone” of the spectrum, where nothing at all is shown. Their proposal would require years of hard work, yielding only a very small chance of providing a reliable answer for a minority of patients, whereas the majority of patients would continue to be treated with unvalidated interventions, based solely on opinion.

Here is where we must address the issue of equipoise. The authors of the letter noted that we did not address this crucial problem. We intentionally sidestepped this complexity, but will here try to briefly summarize our view. To many clinicians, equipoise suggests that only patients for whom the doctor thinks both options are equally balanced
should be enrolled in a trial. This is the best recipe for not enrolling patients; the notion is fragile and unstable (like a ball momentarily in equilibrium on the top of the hill [Fig. 2 upper]). Physicians or patients almost always have some hunch or preference that tips the ball in one direction or another. With this concept of equipoise, evidence supporting one choice or the other can never be obtained.

Freedman, who introduced the notion, attempted to correct this mistake almost 30 years ago by introducing “clinical” or “community” equipoise. Equipoise was initially proposed to justify trial participation, as a solution to the “central problem of research ethics,” recently summarized as follows: “How can physicians offer their patients optimal medical care at the same time that their treatment is selected by chance in the context of a randomized clinical trial?”

The confusion comes from the prescientific idea that by being proposed for trial participation as “research subjects,” patients are being denied the care they need. Optimal medical care presupposes at least 2 options, a comparison between the options, and a proper way to make valid comparisons. Of course, this is the scientific method integral to trials and trials alone, without which “optimal care” does not mean anything at all. Once it is understood that, given the uncertainty, no one really knows prior to the end of the trial in which arm the “optimal care” is located, the problem should dissolve.

Whether we are looking at beliefs, actions, or evidentiary justification for actions, this notion of equipoise misses the crucial difference between the 2 options. What is certain is that an active intervention means an immediate risk. The evidentiary bar for choosing conservative management, which minimizes immediate risks (as ARUBA showed) is not as high (Fig. 2 lower). To practice outcome-based interventions, we need a much more stable notion of uncertainty, one that cannot be so easily overturned with good intentions and variable opinions. The ball is comfortably sitting in the valley of uncertainty, which calls for hard work to bring it up to the evidentiary bar to justify a risky intervention. Invasive interventions that have yet to show their clinical benefits cannot be prescribed in the same way as validated interventions. Doctors can propose interventions, but they must admit that benefits have yet to be proven and that there is a chance that intervening may be wrong. This, basically, is a trial.

What is needed is the notion that unvalidated interventions cannot be prescribed and practiced the same way that validated or normal care can be recommended and practiced. Although it is hard for doctors and patients to understand, what should be emphasized when trial participation is proposed is not that treatments are allocated by chance, but that participation in the trial means that risks are now balanced for the benefit of the very patient being proposed for the trial: patients are offered a 50% chance of getting a promising intervention, hypothesized to lead to better long-term outcomes, and an equal 50% chance of being exempted from the immediate risks of an intervention that may turn out to be harmful. Once it is understood that, in the meantime, optimal care is provided within the trial (in stark contrast to treatments administered outside of trials, which are questionable, unverifiable care forever), there is no need for a notion of equipoise to justify the trial.

This theoretical sketch of what needs to be done recognizes that some clinicians and patients may still opt out
Finally, we draw encouragement from Drs. Elhammady and Heros’ statement that “there has been a strong demand from government and health insurance agencies for evidence-based medicine and outcome measures. The burden of proof has fallen on our shoulders to prove the effectiveness of our interventions.” Of course, that burden has always been on our shoulders, but it is high time that we acknowledge the responsibility. The trials we propose are not designed to justify medical costs, or to meet the demand of government or health insurance agencies, although the results of properly designed trials will accomplish these secondary goals. The primary goal remains to provide patients with optimal care in the presence of uncertainty.

References