A Perfect Storm: How A Randomized Trial of Unruptured Brain Arteriovenous Malformations' (ARUBA's) Trial Design Challenges Notions of External Validity

Kevin M. Cockroft, Mahesh V. Jayaraman, Sepideh Amin-Hanjani, Colin P. Derdeyn, Cameron G. McDougall and John A. Wilson

Stroke. 2012;43:1979-1981; originally published online June 5, 2012; doi: 10.1161/STROKEAHA.112.652032

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/43/7/1979
A Perfect Storm

How A Randomized Trial of Unruptured Brain Arteriovenous Malformations’ (ARUBA’s) Trial Design Challenges Notions of External Validity

Kevin M. Cockroft, MD, MSc; Mahesh V. Jayaraman, MD; Sepideh Amin-Hanjani, MD; Colin P. Derdeyn, MD; Cameron G. McDougall, MD; John A. Wilson, MD

The management of unruptured brain arteriovenous malformations (BAVMs) is controversial and uncertainties exist as to how best to care for these patients. The A Randomized Trial of Unruptured Brain Arteriovenous Malformations (ARUBA)\(^1\) study attempts to shed light on some of these issues. However, the complexity of the disease process, the considerable variation in treatment options, and the trial’s actual design threaten its external validity making it unlikely that significant useful information will be obtained. The following comments are the opinion of the Society of Neuro-Interventional Surgery as well as the Cerebrovascular Section of the American Association of Neurological Surgeons and Congress of Neurological Surgeons. The authors have no financial interest in the ARUBA trial.

As many who care for patients with BAVMs already know, the ARUBA trial seeks to determine whether or not the risks of treatment outweigh the risks of conservative management at 5 years for patients with unruptured BAVMs. Unfortunately, despite recent changes to the study, the trial remains significantly flawed. The trial continues to be plagued by concerns over inconsistent equipoise, overall structure, and selection bias. Additionally, the time horizon of the trial leads to a limited duration of follow-up, which is challenging for a disease with a lifelong threat of death and disability. All of these factors come together to create a trial whose final outcome will have limited external validity.

Treatment of asymptomatic BAVMs clearly presents a significant clinical dilemma. The decision to treat a patient with an asymptomatic BAVM, and if so how to treat that patient, is necessarily based on a variety of considerations. The patient’s clinical situation and the natural history of the lesion are perhaps the most important considerations when deciding whether to treat. The exact method of treatment should be planned after a careful analysis of the BAVM’s radiographic characteristics. The overall goal must be to treat only the patient/BAVM in which the perceived risks of treatment are less than the presumed natural history risk.

The medical literature is replete with studies documenting a variety of different natural history profiles and treatment risks associated with BAVMs. Such a diversity of literature certainly leads to many legitimate differences of opinion regarding BAVM management. However, a difference in opinion as to how to manage a patient with BAVM does not necessarily imply equipoise.

Over the years it has become clear to many that randomized controlled trials cannot answer all the questions in clinical medicine.\(^2\)–\(^4\) In particular, where procedures are involved, trial design severely impacts a trial’s final outcome. ARUBA is no exception. Unlike other surgical trials such as the North American Symptomatic Carotid Endarterectomy Trial (NASCET),\(^5\) the Carotid Revascularization Endarterectomy Versus Stenting Trial (CREST),\(^6\) and Stenting and Aggressive Medical Management for Preventing Recurrent Stroke and Intracranial Stenosis (SAMMPRIS),\(^7\) ARUBA does not require any specific facility or physician accreditation to participate. Moreover, the criteria for enrollment are exceedingly broad. Essentially, any adult patient with an unruptured (not necessarily asymptomatic) BAVM can be enrolled so long as the evaluating physician believes that treatment to eradication is possible and that the physician is uncertain about the best management method. Although catheter arteriography remains the standard for the diagnosis of BAVM, patients may be enrolled in ARUBA on the basis of CT or MRI alone. The trial’s primary outcome is assessed at only 5 years and the evaluation of outcome is determined by physicians who are not blinded to the patient’s treatment arm. Unlike many large trials of this type, ARUBA does not require detailed screening logs, so it is not possible to understand the overall clinical population from which the enrollments are made or the method used to manage those who are not enrolled. Although this trial design cannot be characterized as invalid, these design choices will predictably
influence outcome. In the case of ARUBA, all these choices serve to make it extremely unlikely that a benefit for intervention will be found.

Selection bias is particularly problematic for a prospective randomized controlled trial such as ARUBA. The strengths and weaknesses of various treatments will vary across institutions. Although not necessarily correct, many busy centers have well worked out algorithms for BAVM treatment based on their perceived capabilities. These varied but established treatment patterns at high-volume centers tend to limit the number of patients such centers are willing to enroll in a clinical trial. Patients with BAVM with features associated with rupture and patients with perceived low risk for treatment will likely be offered treatment at these sites. Only patients for whom the perceived risk/benefit for intervention is equivocal would be considered for inclusion in this study. The result, unfortunately, will be low enrollment from these centers and a heterogeneous group of nonrepresentative BAVMs.

The results of this are already evident with enrollment in ARUBA being slower than the investigators anticipated. From the trial initiation in April 2007 to mid-June 2010, only 124 patients were randomized despite over 900 patients screened. Of the 74 sites listed as participating on the National Institutes of Health’s clinical trials web site (http://clinicaltrials.gov/ct2/show/study/NCT00389181), 8 centers from the United States have withdrawn for various reasons and the majority of active sites are now located outside of North America (38 in Europe, Australia, and Brazil). Given the volume of patients with BAVM seen at many of these centers and the number of patients actually recruited, strong selection bias remains a very real concern. Selection bias may be a significant problem at busy US centers. For example, at the University of California at San Francisco, a participant in ARUBA since its inception, and a site where >50 new unbled BAVM patients are seen per year, only 4 patients have been enrolled to date. Unfortunately, as previously mentioned, the lack of detailed screening logs prevents a reasonable understanding of this bias. Nevertheless, it may be possible to draw a parallel to the work of the International Study of Unruptured Intracranial Aneurysms (ISUIA) group. ISUIA was essentially a registry of patients who either declined or were not offered treatment for an unruptured aneurysm. Most physicians enrolled patients they felt were at low risk for hemorrhage from their aneurysm, and this may have resulted in the very low annual rate of rupture for some small aneurysms that was observed in the trial. In ARUBA, clinicians are faced with a similar dilemma and may again select to enroll only those patients whose perceived risk of an adverse event from their BAVM is low.

If the preponderance of enrollments ends up being from low-volume centers, this creates an additional area of concern. Many studies have documented a relationship between volume and outcome in cerebrovascular disease and it is unlikely that the management of BAVMs will be any different. Indeed, in 2010, ARUBA’s investigators reduced planned enrollment from 800 patients to only 400 patients. This change results in a lower power to detect a significant difference between groups and makes a larger difference between groups necessary to reach statistical significance. Such a change would only be reasonable if event rates were higher than initially anticipated when the study was designed. Thus, the potential exists that ARUBA will consist of many BAVMs with a low natural history risk enrolled from predominantly low-volume centers with higher than typical complication rates.

Using parallels from the carotid endarterectomy clinical trial experience, it may be argued that treatment outcomes at high-volume centers do not reflect “real-world” results. However, this is an inappropriate comparison. Carotid endarterectomy is one of the most commonly performed procedures for the prevention of ischemic stroke, a very common disease process, and a high volume of carotid endarterectomies are performed in community hospital settings. This is the opposite of BAVM, a rare disease in which the majority of treatments are performed at tertiary care hospitals with experience and expertise in the management of various cerebrovascular disease states. When patients from such centers are not well represented in a clinical trial of this type, then it is the real-world experience for BAVM treatment that is not well represented.

Trial duration and the period of clinical follow-up warrant further discussion. Primarily because of the structure of National Institutes of Health grant funding, a 5-year end point is typical of prospective randomized controlled trials like ARUBA. However, the 5-year time horizon is a short time period and has little relevance in the clinical decision-making for BAVMs. BAVMs are a congenital disease that typically do not present until the patient is aged 20 to 30 years. In addition, the incidence of clinical presentation is not static over the patient’s lifetime. The information gathered from 20 years of follow-up in 200 patients is vastly different from the information obtained from 5-year follow-up in 800 patients, although the number of patient-years is the same. Put another way, even the most aggressive surgeon or interventionalist would be unlikely to treat a patient with an unruptured BAVM with a life expectancy of only 5 years. The trial investigators do plan to extend follow-up to 10 years, but even 10 years means little to a patient in their 30s with an average life expectancy of 40+ years.

Unfortunately, the culmination of all these factors is a perfect storm that severely limits the external validity of ARUBA. External validity is arguably the most important aspect of any clinical trial. The idea that the results from the study of a subset of patients with a given disease can be extrapolated to the entire disease population, or at least to a much larger subset of that population, is the reason millions of research dollars are spent. However, for such translation to occur, patients enrolled and studied must be representative of the population to which the study’s conclusions will be applied. Given the selection bias inherent in ARUBA, it is unlikely that the general population of BAVM will be adequately represented in the study. Therefore, it will not be possible to extrapolate ARUBA’s conclusions to the management of all BAVM. The alternative then is that ARUBA’s results will be applicable to a specific subset or subsets of BAVM. This too seems unlikely. With such broad enrollment criteria and heterogeneous clinician/center experience, it will
be difficult to achieve sufficient numbers in any given subset to make generalization appropriate. The study is powered for overall comparison of treatment versus no treatment; looking for relevance among specific subsets through post hoc analyses is notoriously problematic.

In the end, the question is not whether uncertainty exists around the management of unruptured BAVM. The authors of this commentary clearly stipulate that uncertainty does exist. The issue is whether ARUBA will actually help to resolve this uncertainty in a significant proportion of patients. Due to the constraints mentioned, it appears unlikely that any widely generalizable information will be obtained. Not all prospective randomized controlled trials constitute Level I evidence and to say that a prospective randomized controlled trial is the only good scientific method to study a disease process is both narrow-minded and shortsighted. A rare disease, with the complexity of BAVM, the wide variety of treatment options available and the prolonged natural history, is probably much better suited to study with a long-term multicenter, consecutive patient registry. Such a registry would need to have third-party adjudication of enrollments and clinical outcomes. This would certainly be an expensive national endeavor but one which would likely provide a greater wealth of management data on all types of BAVM than the present ARUBA trial. In the meantime, however, a significant number of federal dollars will continue to be spent on a trial whose outcome, already a foregone conclusion, has little hope of impacting the management of the majority of patients with BAVM.

Disclosures

None.

References


Key Words: AVM cerebrovascular disease endovascular treatment interventional neuroradiology intracranial hemorrhage methodology neurosurgery randomized controlled trials