Responses to ARUBA: a systematic review and critical analysis for the design of future arteriovenous malformation trials

Elsa Magro, MD,1,2 Jean-Christophe Gentric, MD,1,3 Tim E. Darsaut, MD,4 Daniela Ziegler, MSI,5 Michel W. Bojanowski, MD,1 and Jean Raymond, MD6

1Department of Surgery, Service of Neurosurgery, Centre Hospitalier de l’Université de Montréal (CHUM), Notre-Dame Hospital; 2Direction de l’Enseignement et de l’Académie CHUM|Bibliothèque; and 3Department of Radiology, Service of Neuroradiology, CHUM, Notre-Dame Hospital, Montreal, Quebec; 4Department of Surgery, Division of Neurosurgery, University of Alberta Hospital, MacKenzie Health Sciences Centre, Edmonton, Alberta, Canada; 5Service de Neurochirurgie, CHU Cavale Blanche, INSERM UMR 1101 LaTIM, Brest; and 6Groupe d’étude de la Thrombose en Bretagne Occidentale, Brest, France

OBJECTIVE The ARUBA study (A Randomized Trial of Unruptured Brain Arteriovenous Malformations [AVMs]) on unruptured brain AVMs has been the object of comments and editorials. In the present study the authors aim to systematically review critiques, discuss design issues, and propose a framework for future trials.

METHODS The authors performed a systematic review of the French and English literature on the ARUBA study published between January 2006 and February 2015. The electronic search, including the Cochrane Library, MEDLINE (PubMed and Ovid), CINAHL, and EMBASE databases, was complemented by hand searching and cross-referencing. The comments were categorized as items related to the design, the conduct, and the analysis and interpretation of the trial.

RESULTS Thirty-one articles or letters were identified. The pragmatic design, with heterogeneity of patients and lack of standardization of the treatment arm, were frequently stated concerns. The choice of outcome measures was repeatedly criticized. During the trial, low enrollment rates, selection bias, and premature interruption of enrollment were frequent comments. The short follow-up period, the lack of subgroup analyses, the lack of details on the results of the various treatments, and a contentious interpretation of results were noted at the analysis stage. A fundamental problem was the primary hypothesis testing conservative management. The authors believe that other trials are needed. Future trials could be pragmatic, test interventions stratified at the time of randomization, and look for long-term, hard clinical outcomes in a large number of patients.

CONCLUSIONS In the authors’ view, the ARUBA trial is a turning point in the history of brain AVM management; future trials should aim at integrating trial methodology and clinical care in the presence of uncertainty.

http://thejns.org/doi/abs/10.3171/2015.6.JNS15619

KEY WORDS brain arteriovenous malformation; randomized trial; care trial; A Randomized Multicenter Clinical Trial of Unruptured Brain AVMs; vascular disorders

Brain arteriovenous malformations (AVMs) are a heterogeneous group of rare lesions that can present with headaches, seizures, neurological deficits, and intracranial hemorrhages. Risks of hemorrhage have been estimated in the range of 2%–4% per year.8,15,32,47 Risk factors for future ruptures have been identified, but their use to inform decision making is error-prone, given the methodological problems with AVM observational studies and the modest relative risks.40

Treatments offered may include resection, radiation therapy, and endovascular embolization, alone or combined. Management strategies and their indications have

ABBREVIATIONS ARUBA = A Randomized Trial of Unruptured Brain Arteriovenous Malformations; AVM = arteriovenous malformation; mRS = modified Rankin Scale; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

ACCOMPANYING EDITORIAL See pp 481–485. DOI: 10.3171/2015.7.JNS151408.


INCLUDE WHEN CITING Published online April 28, 2016, DOI: 10.3171/2015.6.JNS15619.

©AANS, 2017
assigned to medical management.30

A decade-long debate over whether preventive eradication of A VMs offers a clinical benefit prompted Mohr et al. to conduct the first trial of unruptured brain A VMs (the ARUBA [A Randomized Trial of Unruptured Brain Arteriovenous Malformations] study), a prospective, multicenter, parallel design, nonblinded, randomized controlled trial, initially involving 104 clinical sites in 9 countries (https://clinicaltrials.gov/ct2/show/NCT00389181).30 The ARUBA study compared patients with an unruptured brain AVM allocated to either medical management alone or medical management and interventional therapy (including surgery, embolization, and radiotherapy, alone or combined). The primary outcome was time to the composite event of death from any cause or symptomatology (including surgery, embolization, and radiotherapy, alone or combined). The primary outcome was time to the composite event of death from any cause or symptomatology alone or combined. An interim analysis performed after 6 years in 223 patients showed that the primary outcome was reached in 35 (30.7%) of the 114 patients assigned to interventional therapy and in 11 (10.1%) of the 109 patients assigned to medical management.30

Premature interruption of the ARUBA trial was followed by numerous editorials and comments. In this study we review critiques of ARUBA and discuss alternative designs. We then propose a framework for future trials.

**Methods**

The systematic review was performed according to guidelines on searching strategies published in the Cochrane Handbook for Systematic Reviews of Interventions (http://handbook.cochrane.org), in the proposal for reporting meta-analysis of observational studies in epidemiology,50 and in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)26 statement.

The literature in French and English published between January 2006 and February 2015 was reviewed. We first identified 10 articles or comments in PubMed (MEDLINE) (E.M.). We collected the specific medical subject headings (MeSH) and key words linked to these articles to start the search strategy (E.M., D.Z.). The search interfaces were: PubMed for MEDLINE, Ovid for MEDLINE, EMBASE, and Cochrane, and EBSCO for CINAHL. We used terms from controlled vocabularies (MeSH for MEDLINE, EMTREE for EMBASE, and CINAHL headings for CINAHL) and words from “all fields” for Cochrane Central, and from “title” and “abstract” in MEDLINE, EMBASE, and CINAHL.

The search strategy and the key words used are available online in the Appendix. Comments extracted by 1 author (E.M.) were reviewed by 2 other authors (J.C.G. and J.R.). Positive comments on the ARUBA trial were summarized. The critiques were categorized as items related to the design, the conduct, and the analysis and interpretation of the trial, keeping in mind that recurring comments referring to the same problems may be explained in different terms. Criticisms and suggestions were further categorized as frequent (mentioned in > 10 articles) or repeated (2–10 articles).

**Results**

The literature search yielded a total of 1556 records. Duplicates were identified and removed; 1166 records were screened; 1063 were excluded because they were not related to ARUBA. One hundred three articles were analyzed in full text: 33 articles from the ARUBA Collaborative Group and 39 articles not containing detailed comments on the ARUBA study were excluded, leaving 31 articles for the present analysis (Fig. 1).

The merit of conducting the first randomized trial regarding an uncommon problem in heterogeneous patients and in difficult conditions was frequently emphasized10,11,13,20,23,25,31,42–44,52 (11 articles). Six articles recognized that the trial was an “important step,” a “warning,” an “alarm,” a “wake-up call,” an “opportunity” to provide evidence regarding the merits of interventions.6,11,13,31,42,49

Four articles acknowledged that the study provided information on the natural history or the risk of treating brain AVMs11,20,44,49

Three articles noted that brain AVMs in the study cohorts were relatively well matched for size, location, and venous drainage pattern, as well as for Spetzler–Martin grade,4,36,49

3 others noted the concordance of ARUBA’s results with previous natural history and population-based studies.1,3,13,31

Concerns and criticisms are summarized in Table 1. The heterogeneity of patients and the lack of standardization of the treatment arm were frequent concerns regarding trial design (> 10 articles). The choice of outcome measures and the primary hypothesis in favor of medical management were repeatedly criticized (9 and 5 articles, respectively). Regarding the conduct of the trial, frequent criticisms were the low enrollment rate (emphasized in 11 articles) and the suspected recruitment bias, including the small number of patients with Grade I–II AVMs treated surgically (> 15 articles). Premature interruption of enrollment was noted in 3 articles.

Concerning the analyses and interpretation of ARUBA, the short follow-up period and inappropriate conclusions were frequent critiques (> 10 articles), and the lack of subgroup analyses (9 articles) and the lack of detail regarding treatment results were repeatedly noted (5 articles). Suggestions included continuing the follow-up of recruited patients over a longer period of time, keeping a registry,3,4,10 and designing other randomized studies.6,23,35

**Discussion**

We discuss the frequent comments and criticisms and propose a framework for future trials.

**Critiques Related to the Pragmatic Design of the Study**

It seems appropriate to begin with a review of the distinctions between explanatory and pragmatic trials, for many comments were related to the pragmatic design of the ARUBA trial.7,10,11,23 Explanatory trials, with narrow selection criteria and tightly controlled protocols reserved
for experts in referral centers, focus on mechanistic end points (AVM obliteration rates for example) and impose protocol-specified tests and visits in addition to those that are clinically indicated (Table 2). An explanatory trial is designed to answer research questions such as “can therapy work in the best circumstances?” If, in spite of optimal conditions for the therapy to succeed, the trial is negative, it is appropriate to conclude that therapy should be abandoned. However, a positive explanatory trial cannot be used to claim that therapy works in the real world, because it was not tested in the real world. Only a pragmatic trial can come to such a conclusion.

Pragmatic trials such as ARUBA accept the diversity of the real world. They should minimally affect normal care. They mainly focus on clinical outcomes when patients are treated under usual conditions. The pragmatic design is the best one to answer clinical questions such as “does therapy work in usual circumstances?” A positive pragmatic trial means it is worth adopting the therapy in practice. A negative pragmatic trial, however, remains inconclusive, because outcomes could have been different under different conditions (selection criteria, operators, or patients).

At the time of trial initiation, all treatment options were already in routine use in ARUBA centers. The question was not whether therapy could work in some patients under ideal circumstances, for that was assumed to be the case. The question was whether preventive therapy should be adopted in normal practice. A pragmatic design was thus appropriate.

Concerns Regarding the Heterogeneity of Patients and the Grouping of Different Management Options

Many authors commented on the heterogeneity of patients, including those with AVMs of various grades, which perhaps calls for different management strategies. A related concern was the aggregation of the various treatment modalities that perhaps preferably should have been tested separately. Heterogeneity can provide inductive strength to pragmatic trial results. A management strategy shown to be superior for a wide variety of patients is more likely to work for future patients than one shown to be effective, but only in a narrow subset of patients.

The alternative would be to design various trials, each specifically addressing a single modality in the treatment of selected patients. This approach spawns a number of problems. Tight selection decreases eligibility. Multiplying the number of narrow-scope trials would probably endanger the feasibility and decrease the clinical relevance of each trial. Depending on the number of patient categories, treatment options, combinations, and comparisons, trials can multiply beyond control. It is unclear how centers would choose to participate in one or several trials, and how patients would be selected to participate in one rather than the other trial. In the absence of standardization of indications (and of evidence) for various management choices, it is unclear which treatment should be favored by design over the others.

Aggregating management options in a single arm (as in ARUBA) can be done. It is sometimes better to compare the experimental treatment A against standard options B, C, or D, particularly if A is innovative (i.e., a new drug or device). When A is shown to be superior to any or all other alternatives, it is likely to work in practice, whereas superiority to only one of the other available options would leave considerable uncertainty regarding the proper clinical adoption of treatment A. That being said, what is peculiar...
about ARUBA is that the strategy that was singled out to be tested against all others was conservative management, a problem that we will discuss below.

When multiple treatment options are grouped into a single arm, results of each treatment modality should be given separately, even though they cannot validly be compared with one another because they have not been randomized.

Concerns Regarding the Choice of Outcome Measures

A frequent comment concerned the choice of the primary end point: death (any cause) or stroke, which was defined as any event (any symptom, including headache or seizure) associated with any imaging finding on CT or MR studies (fresh blood or low-density lesions, for example).3,20,22–25,33,46,49 The protocol mentioned that this

<table>
<thead>
<tr>
<th>Comments &amp; Criticisms (no. of studies)</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inappropriate primary &amp; secondary end points (9)</td>
<td>Bambakidis et al., 2013; Knopman &amp; Stieg, 2014; Lawton &amp; Abla, 2014; Mathiesen, 2008; Meling et al., 2014; Mocco et al., 2013; Pierot et al., 2014; Solomon &amp; Connolly, 2014; Starke et al., 2013</td>
</tr>
<tr>
<td>Heterogeneity of pts; selection criteria (12)</td>
<td>Amin-Hanjani, 2014; Bambakidis et al., 2013, 2014; Cockcroft et al., 2012; Day et al., 2014; Knopman &amp; Stieg, 2014; Mathiesen, 2008; Meling et al., 2014; Mocco et al., 2013; Molina &amp; Selim, 2014; Proust et al., 2014; Russin &amp; Spetzler, 2014</td>
</tr>
<tr>
<td>Lack of standardization of treatment arm (14)</td>
<td>Amin-Hanjani, 2014; Bervini et al., 2014; Cockcroft et al., 2012; Day et al., 2014; Grasso, 2014; Heros, 2014; Knopman &amp; Stieg, 2014; Mathiesen, 2008; Meling et al., 2014; Mocco et al., 2013; Pierot et al., 2014; Russin &amp; Spetzler, 2014; Solomon &amp; Connolly, 2014; Yen et al., 2014</td>
</tr>
<tr>
<td>Design &amp; primary hypothesis in favor of medical management (5)</td>
<td>Bojanowski et al., 2015; Cockcroft et al., 2012; Mathiesen, 2008; Mocco et al., 2013; Rutledge et al., 2014</td>
</tr>
<tr>
<td>Low enrollment rate (11)</td>
<td>Al-Shahi Salman et al., 2014; Amin-Hanjani, 2014; Bambakidis et al., 2013, 2014; Bervini et al., 2014; Knopman &amp; Stieg, 2014; Meling et al., 2014; Pierot et al., 2014; Proust et al., 2014; Rutledge et al., 2014</td>
</tr>
<tr>
<td>Recruitment bias; included low rate of surgical management for Spetzler-Martin Grade I–II, &amp; high rate of embolization or radiosurgery alone (18)</td>
<td>Amin-Hanjani, 2014; Bambakidis et al., 2013, 2014; Bervini et al., 2014; Cockcroft et al., 2012; Elhammady &amp; Heros, 2014; Gross et al., 2014; Heros, 2014; Knopman &amp; Stieg, 2014; Lawton, 2015; Lawton &amp; Abla, 2014; Meling et al., 2014; Mocco et al., 2013; Molina &amp; Selim, 2014; Proust et al., 2014; Russin &amp; Spetzler, 2014; Rutledge et al., 2014</td>
</tr>
<tr>
<td>Premature interruption of enrollment (3)</td>
<td>Al-Shahi Salman et al., 2014; Bambakidis et al., 2013; Rutledge et al., 2014</td>
</tr>
<tr>
<td>Lack of subgroup analyses (9)</td>
<td>Bambakidis et al., 2013; Day et al., 2014; Lawton, 2015; Meling et al., 2014; Molina &amp; Selim, 2014; Pierot et al., 2014; Russin &amp; Spetzler, 2014; Starke et al., 2013; Yen et al., 2014</td>
</tr>
<tr>
<td>Short follow-up (17)</td>
<td>Amin-Hanjani, 2014; Bambakidis et al., 2013, 2014; Cockcroft et al., 2012; Day et al., 2014; Gross et al, 2014; Heros 2014; Knopman &amp; Stieg, 2014; Meling et al., 2014; Mocco et al., 2013; Pierot et al., 2014; Pollock et al, 2013; Russin &amp; Spetzler, 2014; Rutledge et al., 2014; Starke et al., 2013; Warlow, 2014; Yen et al., 2014</td>
</tr>
<tr>
<td>Lack of details regarding the treatment arms (5)</td>
<td>Grasso, 2014; Lawton, 2015; Lawton &amp; Abla, 2014; Pierot et al., 2014; Russin &amp; Spetzler, 2014</td>
</tr>
<tr>
<td>Inappropriate conclusions (14)</td>
<td>Amin-Hanjani, 2014; Bambakidis et al., 2013, 2014; Day et al., 2014; Elhammady &amp; Heros, 2014; Grasso, 2014; Lawton, 2015; Lawton &amp; Abla, 2014; Meling et al., 2014; Mocco et al., 2013; Pierot et al., 2014; Proust et al., 2014; Russin &amp; Cohen-Gadol, 2014; Russin &amp; Spetzler, 2014; Warlow, 2014</td>
</tr>
<tr>
<td>Continued follow-up (7)</td>
<td>Day et al., 2014; Grasso, 2014; Mohr et al., 2014; Molina &amp; Selim, 2014; Russin &amp; Spetzler, 2014; Warlow, 2014; Yen et al., 2014</td>
</tr>
<tr>
<td>Registry of all pts w/ AVMs (3)</td>
<td>Bambakidis et al., 2013, 2014; Cockcroft et al., 2012</td>
</tr>
<tr>
<td>Beyond ARUBA: Randomized Low-Grade Brain AVM Study, Observation versus Surgery [BARBADOS] (2)</td>
<td>Lawton, 2015; Potts et al., 2015</td>
</tr>
<tr>
<td>Treatment of Brain AVMs [TOBAS] care trial &amp; registry: an RCT (1)</td>
<td>Bojanowski et al., 2015</td>
</tr>
</tbody>
</table>

Pts = patients; RCT = randomized controlled trial.
choice of primary end point would “avert judgment about severity,” while a secondary end point (modified Rankin Scale [mRS] score > 1) concentrated on significant events associated with impairment. The reasoning behind these choices is that most patients with unruptured AVMs are neurologically intact. Any impairment or disability resulting from preventive treatments should count against those treatments. For many surgeons and interventionists, these end points were too “soft,” because patients frequently have headaches after embolization or resections, and traces of blood or low-density lesions of hardly any clinical significance are common on postoperative scans. This impression is reinforced by the data, with 30.7% of treated patients reaching the primary end point. It is unclear from the report, however, how many events were perioperative and how many were permanent deficits, other than for one secondary end point (mRS score > 1), which was found in 46.2% of the 52 patients followed for 30 months.

**Concerns Regarding the Conduct of the Study**

**Slow Recruitment**

Many authors noted the gap between the number of screened patients and those enrolled, and either suspected that selection had taken place, or hypothesized that few (American) major surgical centers participated. Others remarked that few (American) major surgical centers participated, or that few patients were treated with resection alone. It is true that enrollment lagged behind expectations. Many centers enrolled a single or no patient at all. The choice of curative options, left to participating investigators, may not be representative of the decisions made by nonparticipating centers. Importantly, even if selection had occurred, it would not compromise the scientific validity of the comparison made between the randomized groups, although it may decrease the clinical applicability of trial results.

Many of those problems affect virtually all trials, and other than constantly reminding clinicians that they should participate in ongoing trials, there is no easy solution. Ethics and regulations ensure that participation will always be voluntary and subject to clinical judgment.

Until trials become better integrated into practice or unless some incentive is attached to participation, meager enrollment of patients and centers is to be expected.

**Premature Interruption**

The protocol determines a priori stopping rules. Once the interim data show that a predetermined safety boundary has been crossed, what choice remains other than trial interruption? We must be conscious of potential extraneous influences. Trial completion can depend on an arbitrary, prefixed funding framework. Trial feasibility within a 5-year time frame, a common deadline for public funding programs, may have been a concern. In this instance, a difficult, costly, slow-recruiting trial finally provided data compatible with a conclusion.

**Outcome Analyses**

The primary hypothesis of the ARUBA protocol was that medical management would improve the long-term outcome as compared with interventional management. The long-term notion is important to patients and clinicians. However, the primary outcome measure was time to (first) stroke or death, stroke being defined as any symptom associated with any imaging finding. It seems that we have lost the notion of “long term,” and the definition used for “stroke” appears to be too broadly inclusive to qualify as a long-term outcome. It is true that a secondary outcome was proposed to look at impairment at 5 years (mRS score > 1), but this end point was limited to a fraction of the patients (39%) at 36 months.

The primary outcome was studied using the Kaplan-Meier method and measured using Cox proportional-hazards regression models. Although the authors reported “no departure from the proportionality assumption” when it was tested post hoc, such testing may not have had the power to identify what was obvious a priori. For each treated patient there are 2 parts to this trial: a treatment period and a follow-up period. By intervening preventive, risks of stroke or death in the treatment period can only be increased, as compared with medical management, followed, when successful, by an inversion of the ratio, with a relative decrease of stroke or death from ruptures dur-
ing the follow-up period. Analyses published in the Lancet were performed within the treatment period of 73 of the 114 patients (“at the time of analysis, 53 patients randomized to interventional therapy had ongoing treatment plans, whereas 20 had not yet initiated therapy”). The fundamental assumption of a Cox model (a constant hazard ratio between the 2 groups) is clearly not met here.

This may seem to be a minor point, but the mismatch between the initial goals of the trial and the final methods used to measure outcomes reveals an all-too-common problem that will be discussed below: statistical and trial methods are disconnected from what patients and clinicians really need.

**Interpretation of Findings**

Many commentators have contested the conclusions of the study. The Lancet interpretation is that “after 33 months of follow-up, the ARUBA data provide evidence that medical management alone is superior to preventive interventional management for the prevention of death or stroke in patients diagnosed with an unruptured brain AVM.” This interpretation, with a minor modification, is perhaps uncontestable: after 33 months of follow-up, medical management alone is superior to interventional management for the prevention of treatment-related death or stroke.

The ARUBA study was a daunting endeavor and what it revealed is still a wake-up call. We must thank the investigators for having had the courage to ask a difficult question, for designing an innovative protocol applicable to many centers, for obtaining financial support, for persevering in the face of resistance and slow recruitment, and for providing the community with the first results of a randomized trial on this uncommon condition. We must nevertheless critically analyze where things went wrong and search for a way to come out of the impasse that we now face.

**The Problem of ARUBA and a Potential Solution**

Before we examine the design of future trials, we must review a problem that illustrates why we must learn to better integrate clinical research and care: the wrong-sided ARUBA trial hypothesis. The trial hypothesized that conservative management improved patient outcomes as compared with interventions. This is unusual. Normally, therapy is the experiment that needs to be tested, conservative management being the control group. There are interrelated logical, clinical, and ethical reasons for this order.

It may make sense for a clinician to reveal to a patient something like, “I believe the intervention is promising: it can increase your chances of having a good outcome in the long term, but that remains an unknown hypothesis. I may be wrong. There are immediate risks. The safest way to manage this uncertainty is to participate in a trial that will give you a 50% chance of getting the promising intervention, and a 50% chance of not getting an intervention that could in the end be revealed as harmful. If one day we prove the intervention beneficial, then I will offer some patients a 100% chance of getting the intervention.” If the patient is not interested in trying the intervention, he or she gets conservative management. That is prudent. However, reversing the burden of the proof (to say “I believe the intervention is harmful, but let us prove it in a trial”) is not clinically coherent.

The reason why treatments must be tested (and not conservative management) is that the following is certain: by intervening, risks are taken upfront while benefits (if there are any) remain potential, yet unproven. The burden of the proof is on intervening. For any trial, management options can generally be ordered on a spectrum from less risky, and perhaps less effective, to more risky, more effective treatments, and the order that determines which option needs to be tested and the outcomes verified cannot be reversed. In the absence of evidence of superiority of any treatment of AVMs, the safest approach immediately is conservative. If that approach fails, patients can always be offered the alternative option that they were denied at the time of randomization (the hypothetically more effective treatment), whereas treatments and their inherent risks cannot be undone to offer conservative management.

This hierarchy of management options impacts other aspects of trial design. A pragmatic trial can aggregate various options into one arm of the comparison, but the consequence is that conclusive results are possible only for the other arm of the comparison. A single intervention (A) can be shown to be better than a number of alternatives (B or C), but the lumped interventions cannot be validated as beneficial; otherwise any intervention, even a harmful one (say C), could be validated by piggybacking on a beneficial one (B). Treatments must be individually validated as beneficial on their own merits. By grouping all management alternatives on the other side of the comparison, without stratification, the ARUBA trial could only test and verify that medical management was best. If one argues that the side of the hypothesis does not matter, since statistical testing will be 2-sided, and that medical management was the “experimental arm” in ARUBA, this is strictly speaking not true, because an experimental management option cannot be tested when it is on both sides of the comparison (patients allocated to AVM treatment were also going to receive optimal medical management in addition to the intervention).

Where did things go awry? The clinical need was to try to see if taking upfront risks by intervening to prevent ruptures was worth the potentially improved long-term outcomes. We ended up testing the single option that in the absence of evidence did not need to be tested. Combined with a choice of analyses that could not measure the promised long-term outcomes, the trial was, in a sense, unfair. Furthermore, there is no need for evidence to offer what patients should get no matter what: when trials fail to show superiority of interventions, conservative management wins, until one day one treatment is shown to be beneficial. This is how medical care can progress.

Normally, there is no need for evidence to not offer treatments that have not shown superiority as compared with conservative management. It seems as though ARUBA was needed to protect patients from interventions. All this would not have been possible had our field not adopted treatments without requiring a proof of benefit.

If ARUBA’s results confirm the thesis that intervening in unruptured brain AVMs is experimenting with unprov-
The long-term goal is still to provide an answer as to which urge us to conceive of clinical care research differently. If better understand how trial methodology and care in the management of unruptured brain AVMs. We need to up and cross-overs may increase with time, to name just pragmatic trials are more vulnerable to postrandomization large number of patients with an uncommon disease. Many multicenter trial involving the long-term follow-up of a room to show the potential benefits of therapy. and truly measured in the long term, to provide enough large number of patients must be followed for long time periods to reach a convincing answer. The design can be pragmatic, to minimally affect care, but treatments must be stratified. The primary end point should be a hard outcome and truly measured in the long term, to provide enough room to show the potential benefits of therapy. There are major practical problems in conducting a multicenter trial involving the long-term follow-up of a large number of patients with an uncommon disease. Many funding agencies only support 5-year projects; long-term pragmatic trials are more vulnerable to postrandomization confounding and selection bias; and losses to follow-up and cross-overs may increase with time, to name just a few. These problems are not insurmountable, but they urge us to conceive of clinical care research differently. If the long-term goal is still to provide an answer as to which treatment option is best, in the meantime care trials can provide patients with a chance to access innovative or unproven treatments, and a clinical context for clinicians to practice verifiable care in the interest of current patients. We must learn to better integrate clinical research and care because optimal care must take uncertainty into account.

Conclusions

The ARUBA trial was a turning point in the history of the management of unruptured brain AVMs. We need to better understand how trial methodology and care in the presence of uncertainty can be combined to offer optimal medical care in real time.

References

2. Amin-Hanjani S: ARUBA results are not applicable to all patients with arteriovenous malformation. Stroke 45:1539–1540, 2014

### TABLE 3. Contrast between ARUBA and future trials

<table>
<thead>
<tr>
<th>Feature</th>
<th>ARUBA</th>
<th>Future Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary hypothesis</td>
<td>Testing medical management</td>
<td>Testing interventions</td>
</tr>
<tr>
<td>Standard treatment</td>
<td>Interventions</td>
<td>Medical management</td>
</tr>
<tr>
<td>Experiment</td>
<td>Medical management</td>
<td>Interventions (stratified)</td>
</tr>
<tr>
<td>Primary outcome</td>
<td>Time to any stroke</td>
<td>mRS score &gt;2 at 5 or 10 yrs</td>
</tr>
<tr>
<td>Selection criteria</td>
<td>All-inclusive</td>
<td>Any pt considered for intervention</td>
</tr>
<tr>
<td>Design</td>
<td>Pragmatic</td>
<td>Pragmatic</td>
</tr>
<tr>
<td>Conduct</td>
<td>Interruption of trial when observation shown to be convincingly superior according to a group sequential monitoring procedure</td>
<td>Interruption of trial when 1 treatment is shown to be superior, according to a predefined no. of pts observed over a predefined no. of yrs</td>
</tr>
</tbody>
</table>
Author Contributions
Conception and design: Raymond, Magro, Gentic, Darsaut, Bojanowski. Acquisition of data: all authors. Analysis and interpretation of data: Raymond, Magro, Gentic, Darsaut, Bojanowski. Drafting the article: Raymond, Magro, Gentic, Darsaut, Bojanowski. Critically revising the article: Raymond, Magro, Gentic, Darsaut, Bojanowski. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Raymond. Statistical analysis: Magro, Gentic, Darsaut. Administrative/technical/material support: Ziegler. Study supervision: Raymond, Magro, Gentic, Darsaut, Bojanowski. Literature search and review: Ziegler.

Supplemental Information
Online-Only Content
Supplemental material is available with the online version of the article.

Correspondence
Jean Raymond, Interventional Neuroradiology, CHUM—Notre-Dame Hospital, 1560 Sherbrooke East, Pavillion Simard, Rm. Z12909, Montreal, QC H2L 4M1, Canada. email: jean.raymond@umontreal.ca.