REVIEW

Trial on endovascular treatment of unruptured aneurysms (TEAM): study monitoring and rationale for trial interruption or continuation

Étude sur le traitement endovasculaire des anévrismes non rompus (TEAM) : surveillance et principes d’interruption ou de poursuite de l’étude

J. Raymonda,*, J.F. Mederb, A.J. Molyneuxc, A.J. Foxd, S.C. Johnstoned, J.P. Colletf, I. Rouleaug,
Trial on Endovascular Aneurysm Management (TEAM) collaborative group

a Department of Radiology, Interventional Neuroradiology Research Unit (INRU), Université de Montréal, CHUM Notre-Dame Hospital, 1560 Sherbrooke East, suite M-8203, H2L 4M1, Montreal, Quebec, Canada
b Département d’imagerie morphologique et fonctionnelle, centre hospitalier Sainte-Anne, Paris, France
c Nuffield Department of Surgery, Neurovascular Research Unit, University of Oxford, Radcliffe Infirmary, Oxford, United Kingdom
d Department of Medical Imaging, Sunnybrook and Women’s College Health Sciences Centre, Ontario, Canada
e Department of Neurology, University of California, San Francisco, USA
f Centre for Healthcare Innovation and Improvement, University of British Columbia, Vancouver, Canada
g Département de psychologie, centre de neurosciences de la cognition, université du Québec à Montréal et service de neuropsychologie, CHUM Notre-Dame Hospital, Montréal, Canada

Keywords
Aneurysms; Intracranial; Therapy; Clinical trial; Trial monitoring

Abstract Preventive treatment of unruptured intracranial aneurysms is often performed but has never been proved beneficial as compared to conservative management. In a context of uncertainty, the ‘best treatment’ that can be offered to each individual is a chance to be treated and thus to be protected from rupture of the aneurysm, and an equal chance not to be treated, and hence to be exempted from possible immediate complications, using randomization. Such action is optimal unless or until an independent committee with privileged access to data judges that, given the comparative outcome of the 2 groups, preventive treatment or conservative management, is generally warranted. Potential reasons to interrupt such a study are reviewed, including insufficient recruitment, poor compliance, excessive cross-overs, unacceptable iatrogenia, and treatments being convincingly different or equivalent. We conclude that insufficient recruitment is the sole realistic event that could lead to premature
Clinical studies are designed to provide a valid answer to a significant clinical problem. The problem we wish to address here could be formulated as: ‘Should asymptomatic intracranial aneurysms be treated by endovascular means?’ This question is a clinical problem because preventive coiling of unruptured aneurysms is commonly performed in most endovascular centers while the available ‘evidence’ suggests that the hemorrhagic risks entailed by the presence of the lesion rarely justify the risks of intervention (ISUIA) [11]. The only rational way to conceive that physicians, committed to a ‘therapeutic obligation’ (that for the sake of this article we will capture as ‘always treat patients in a way that will optimize their outcome’) would persist in acting contrary to apparent evidence is that they do not trust the available data. Now conceding that they may be right in rejecting data from biased registries as unreliable [6], one may question what could be this more reliable knowledge they are relying on to justify risky treatment of healthy individuals. We will discard up front extraneous motives such as financial or corporate interests, and thrust aside the suggestion that coiling may be the best approach to palliate an irrational fear of the disease, and concentrate on the hypothesis that such physicians that continue to treat unruptured aneurysms do so because they believe that it may be in the patients’ best interest. Now could their beliefs be solely based on erroneous assumptions, or their actions justified solely on good but ill-advised intentions? Confronted with this conflict between their beliefs and available evidence some physicians are willing to consider that their opinion is fallible, that in systematically taking risks by offering treatment to asymptomatic individuals for a potential but unproven benefit they may be doing more harm than good, and that vociferous opinionated discourses won’t suffice to settle the matter.

The main premise of this worldwide effort (TEAM) is that the strenuous but rational and rigorous manner to settle the matter is to accept the current uncertainty, to submit our beliefs to the verdict of experience, and to search for more reliable evidence using established methods grounded on reality. In so doing, the morally serious physician acts in a manner that respects veracity (regarding the uncertainty), patient autonomy (in acknowledging the uncertainty, allowing choices and providing all reliable information in the informed consent procedure), and maximizes beneficence and minimizes maleficence by offering randomization to two management schemes: preventive coiling or indefinite deferral of treatment.

Clinical studies need to be monitored. Most pressing reasons to monitor trials are to minimize exposure to unacceptable harm, to assure that participants are not denied a beneficial intervention, and to evaluate if study objectives can be reached. It is part of our moral commitment to the subjects submitted to the randomization process that we offer treatment to subjects of the deferred treatment group as soon as, but not before, there is convincing evidence that treatment is beneficial. Other ethical requirements for the randomization process are: 1/ there is uncertainty regarding the best management of the condition and 2/ the trial shows promises of providing a valid answer to the research question. We may never be in a position to start a trial with the certitude that it will be a success, but once an objective evaluation of the progress concludes that the trial will not, or has very little chance of reaching its objectives, it may no longer be judged ethical to continue recruitment. Finally, resources are limited and it would be desirable to stop a trial that cannot provide useful knowledge to reinvest the resources into more promising research.

The investigators, Executive and Steering committees are committed to promote the trial and cannot make this objective evaluation. Hence decisions regarding continuation or interruption are more appropriately entrusted to an independent DSMC, empowered with privileged access to ongoing data.

One central task of the DSMC (Data Safety Monitoring Committee) is to determine whether a study should be terminated prematurely. Reviewing reasons to continue or
interrupt a study may provide another perspective from which to judge the necessity for initiating such a difficult but crucial endeavour. We will first summarize the rational and goals of the trial. Then we will attempt to defuse the apparent dilemma concerning scientific and clinical duties simultaneously involved in performing this clinical trial. Finally we will review in detail reasons for stopping or not the trial in an effort to reach a deeper understanding of the issues at stake.

The study

The management of patients with unruptured aneurysms remains controversial. Patients with unruptured aneurysms may suffer intracranial hemorrhage, but the incidence of this event is still debated. Endovascular treatment can prevent rupture, but involves immediate risks. Furthermore, successful treatment does not eliminate all risks. The safety and efficacy of endovascular treatment of unruptured intracranial aneurysms remain undetermined. Hence the balance of the risks and benefits is uncertain. We propose an international, randomized, multi-centre, controlled study comparing the combined mortality and morbidity (MRS ≥ 3) from intracranial hemorrhage in patients with unruptured aneurysms treated by conservative management (or deferral for 10 years or until definite indications are thought to have arisen) as compared to endovascular coiling.

The entire study will enrol approximately 2002 patients equally divided between the two groups, a size sufficient to achieve 80% power at a 0.0167 significance to detect differences in 1) disease or treatment-related poor outcomes from 7-9% to 3-5%; 2) overall mortality from 16 to 11%. The duration forecast of the study is 14 years, the first three years being for patient recruitment plus a minimum of 10 years of follow-up. The evolution of the trial will be monitored by an independent Data Safety and Monitoring Committee, with interruption in case of excessive treatment-related morbidity, excessive hemorrhages in the conservative group, insufficient recruitment rates or futility, and of course, if one group is convincingly doing better.

The participant: ‘my patient’

A common mistake is to create 2 different worlds with specific methods and values: the clinical world of patients and the scientific world of research subjects. Although psychological and epistemological difficulties with randomization are expected, and perhaps with a frequency that may jeopardize the feasibility of the study, we believe there is no real ethical tension between benefits for participants and future patients. There is only one world of patients with aneurysms, and therefore there can only be one ethical realm. In constructing the rationale of the study, we deny that we attempt to bend our clinical obligations to current patients in order to gain scientific knowledge for future patients. The research question concerns first and foremost our current patients with unruptured aneurysms, for whom no action has yet been proved beneficial.

The trial is a pragmatic one, and except for questionnaires to evaluate outcomes, it does not require any test outside normal practice. Interventions are not experimen-tal, and the procedures have been practiced and standardized for more than a decade. Our aim is to offer to current patients the maximum chances of an optimal outcome, and each participant is enrolled for his own benefit.

In this clinical trial there is no gulf between practical actions and scientific knowledge. Science here is just the appropriate method to minimize error, both in the clinical management of patients and in the quest for the reliable knowledge needed to guide our actions. And our actions are dictated by our therapeutic obligation. It is simply impossible to do what is best for each patient until we know what ‘best’ means. There is here no higher sphere of intellectual knowledge that would justify the sacrifice of a human life. The knowledge we are looking for must be as reliable as possible, and most importantly, it must be practical knowledge gathered from the clinical world of everyday practice, not scientific laws in the tradition of natural philosophy. This practical knowledge must be given the reliability that it deserves, no more and no less. Until it can be asserted with confidence that patients need to be treated, because they do better with treatment than without, the ‘best’ we can offer to our patients is a chance to be treated and thus to be protected from rupture of the aneurysm, and an equal chance not to be treated, and hence to be immune from immediate surgical complications. Hence, when the uncertainty dominates, we offer randomization until the uncertainty can be replaced by reliable evidence. Now how and when the reliability of evidence will be determined is a crucial question that we will only partly cover in the following discussion. We will simply mention for now that this cannot be asserted by individual clinicians facing individual patients, and that determination of the acquisition of reliable knowledge necessitates a communion of data from multiple individuals, and a judgement to be delegated to an impartial and independent group of persons, the DSMC. The peculiarity of the trial, in which clinical care has to be provided in a context of uncertainty, and in which the collection of valid data is done simultaneously, simply reflect the fact that we need ‘to repair the boat while it is at sea’ [3].

Although the primary duty of the DSMC is protection of trial participants, the judgement of the DSMC regarding trial interruption will affect individuals with aneurysms, indiscriminately, whether they are current participants or future patients, in the sense that the committee will have judged that we now know, based on best possible evidence, that patients should be offered treatment (or not). At least in the logic of this clinical research, if they judge that the trial should be interrupted, then the ‘next potential participant’ becomes the first ‘future patient’ and conversely, if it should be continued, the next ‘future patient’ becomes the ‘next potential participant’. If and once the verdict of the DSMC is in favour of coiling, the patient (within the trial) being observed will be offered treatment, and so will be the future patient. Conversely, and all things being otherwise equal, if treatment leads to convincingly worse outcomes, the trial will be interrupted, and a potential candidate for the trial will be offered conservative management, just as any future patient. Some have argued that a ‘policy decision’ requires a greater amount of evidence than the present patient decision [2] and this has been
posed as a moral dilemma in clinical trials, when there may be some evidence in favour of one group, but not sufficient evidence for reaching ‘statistical significance’. But there are numerous occasions when, had trials been stopped early we would have wrongly concluded that treatment is beneficial [9]. In addition, in the face of uncertainty, one may use error-prone heuristics that, though they may intend to yield an optimal outcome for the present patient, systematically lead to worse outcomes, but without any means of detecting this error [7]. Hence, this problem presents itself as a psychological tension, and not a moral dilemma. The care of the present patient can only be optimized by using knowledge that is as reliable as possible, and it is the responsibility of the DSMC to judge when this has been attained.

Principles for study termination

The DSMC is fully independent and the following principles are only guidelines for reflection. Reasons for study termination can be classified as data-dependent or data-independent.

Data-independent stopping reasons may include:

- emerging information, privileged or public, that make the trial irrelevant, unnecessary or unethical. For example, although improbable, the DSMC could find during the course of the trial that another treatment, not included in the trial, is scientifically proven beneficial to trial subjects;
- poor execution or poor quality compromising the ability of the study to meet its objectives, such as inadequate monitoring of adverse events, negligence, misconduct, fraud, or suspicion of conflicts of interest.

In the present study we expect adverse events in the ‘treatment’ group to occur immediately at the time of intervention, during the recruitment years, while they may be compensated by hemorrhagic events in the ‘deferred treatment’ group, which will occur throughout the study that includes a longer, ten-year follow-up period. Thus it may be pertinent here to distinguish a recruitment period and a follow-up period.

Data-dependent stopping reasons include, during the recruitment phase of the study:

- recruitment of centres or enrolment of patients are so slow as to compromise the possibility of the study to meet its objectives;
- compliance to group allocation is so poor that the trial cannot proceed in a meaningful fashion;
- adverse events of treatment are unacceptably frequent or so frequent as to convincingly support the hypothesis that treatment will be shown to cause significant harm as compared to deferred treatment;
- hemorrhagic events in the deferred treatment group are so frequent as to render this option unethical.

Data-dependent stopping reasons include, during the follow-up phase of the study:

- treatments are convincingly different;
- treatments are convincingly not different;
- cross-over is so frequent as to endanger the objectives of the study.

Another way to classify reasons to interrupt a trial is:

- interruption to minimize exposure to unacceptable harm;
- interruption not to deny a beneficial intervention;
- futility.

We will now examine each reason for interruption, starting with futility.

Futility

Relevant to the anticipated difficulties with the present trial is the concept of futility related to inadequate patient enrolment, poor compliance with group assignment and crossover between groups.

Futility because of inadequate recruitment

This type of ‘futility’ focuses on the feasibility of the trial, and it certainly is a real concern. There are multiple resistances confronting this clinical trial. Physicians may think they ‘know’ for each patient what the best option is. Patients may fear the threat of aneurismal rupture and opt for immediate treatment. To go beyond the maxims from immemorial times (such as better be safe than sorry), to question the value of an established treatment, to confront the force of custom and habits, may turn out to be exceedingly difficult.

This concern with feasibility of recruitment may endanger recruitment itself.

There is contradiction in our reasoning and circularity in our behaviour. To point out the contradictions, one often hears that patients are likely to be reluctant to enrol once they are informed of the low risk of treatment. However, one may also hear complaints that ‘the very low risk of hemorrhage will require large numbers and excessively long follow-up’. But if hemorrhage is so rare, and benefits nearly impossible to show, why offer a risky treatment so frequently? When no validation of action is possible, some physicians may find refuge in beliefs and habits, generally in favour of their action. Trials are not feasible because they are considered unfeasible, because equipoise is ‘believed’ to be rare, because we believe risks of treatment/observation are high/low in this/that patient. Unjustified beliefs are strong because there is no evidence and no hope to have any, and the uncertainty is felt by clinicians to be sterile and painful. We need to emerge out of this vicious circle: A repetitive uncertainty is in fact an opportunity for knowledge. Once a physician participates in the trial, and once the trial is ongoing, then he is in a position to
acknowledge the uncertainty. This uncertainty is no longer so painful or sterile, since now there is hope to get evidence. There is a growing feeling that one is frequently uncertain about what to do, and with this rising wave the trial will actually become feasible.

It is our responsibility to attempt to convince the field, potential investigators and participants that the current approach is inescapable.

The failure to achieve sufficient recruitment to reach meaningful results, if and only if the effort was genuine, sustained, optimally organized and sufficiently supported, may however become a fact that would need to be acknowledged by the scientific and clinical community. Even such a disappointing result would be an improvement, since the pessimistic opinion regarding the feasibility of science in this field, remains solely an opinion. In the absence of a real challenge (or of a reasonable try at providing a valid answer), and in the presence of conflicting interests, this opinion is in our view unacceptable. But the failure to recruit a meaningful number of patients within a reasonable time period, given that we gave it a ‘good try’, may have to be accepted as ‘a matter of fact’.

The challenge for the DSMC here is to determine with more precision what could be considered a ‘good’ or even a ‘reasonable try’. The DSMC may wish to fix a minimum of participating centres, and a minimum number of patients at a certain date. We planned recruitment of approximately 1 patient/centre/month, a reasonable goal taking into account an enrolment rate of 1/5 patients with unruptured aneurysms. It is felt essential to the success of the effort to recruit subjects in a ‘compact’ fashion, within 3-4 years. Some delays in the initiation of the centres but a flattening of the accrual rate curve would be much more difficult to accept. One-way of providing a boundary for too slow accrual would be an arbitrary 25% of the expected recruitment rate, or less than 4 patients/centre/year.

Adherence to group assignment

With respect to the concept of compliance, we believe that two different aspects should be distinguished: respect of group allocation at the time of randomization or in the immediate period after randomization and crossover during the 10 years of the follow-up period.

While randomization is the best way to prevent bias, bias can be reintroduced if patients do not comply with their assignment. There are a number of means to promote adherence to group assignment. The patients will be recruited after 2 consultations, to ensure a minimal number of ambivalent candidates and to assure that participants are as comfortable as possible with the concepts of uncertainty and randomization. There may be a certain learning curve in appropriately informing subjects and assuring they can deal with both possible outcomes of the randomization process before enrolment. The subjects allocated to coiling will be treated within one month. The trial team and the participating centres will maintain a competent, compassionate, available environment to respond to questions, discomfort or concerns of subjects allocated to “conservative management” to minimize the potential attraction of active treatments. The DSMC will determine what will be considered acceptable ranges but we expect more than 95% compliance.

Futility because of excessive crossover during follow-up

Once patients have been treated by endovascular coiling, it is of course impossible to ‘crossover’ to ‘conservative management’. Hence there is during the follow-up period only one direction of crossover: the treatment of subjects initially allocated to ‘conservative management’. There are two different causes for so-called crossover that need to be distinguished: patients that will present ‘medical reasons’ for implementing active treatment, such as new neurological deficits related to the aneurysm, intracranial hemorrhages or signs or symptoms of ‘impending rupture’, these patients have in fact reached an endpoint and will be treated as standard medical care dictates. These are not strictly speaking crossovers since treating lesions as definite indications appear is an integral part of conservative management as compared to elective treatment of an asymptomatic lesion.

Other participants that may elect to be treated during follow-up, but without a definite ‘medical reason’, can be categorized as ‘crossovers’ in the sense of an external factor that could affect the meaning of results of the trial and endanger the validity of the conclusions. This type of crossover will be discouraged, within the limits dictated by ethical respect for the dignity and autonomy of participants. The main strategy is one of compassion, availability of participating physicians, information and reassurance (of course when appropriate). Although it is clear that this type of crossover can affect many important aspects of the trial (it would endanger the evaluation of the natural history of the disease, and smudge differences between the 2 groups), another way of judging this result is the collection of human data of extreme importance in this particular problem. Once more we wish to emphasize that priority is given to convince the clinical field, neurointerventional and neurosurgical experts, participating investigators and subjects that we must face the uncertainty regarding the potential benefits versus harm that could result from unproven actions. A dreadful hypothesis regarding the current therapeutic attitude is that we could be using an invasive, potentially risky treatment to cure patients from the fear that we instillate ourselves by the way we approach and explain the meaning of the discovery of the asymptomatic lesion. In other words if treatment benefits are to be calculated in terms of palliating fear, a direct psychological or behavioural approach could accomplish equally beneficial results with much smaller risks and costs. But once we accept that, despite the fact that everything possible was done to approach the problem rationally, an important proportion of patients assigned to conservative management do wish to be treated or succeed in obtaining treatment without strict medical justification during the 10-year follow-up period, we may be forced to accept this result as a significant finding pertinent to the human (as opposed to ‘natural’) history of the disease, and in this important sense it would still be valuable observational data to obtain. Thus we propose that interruption, at least at the time of the follow-up.
phase, because the rate of crossover would be judged ‘excessive’ would be inappropriate.

**Futility defined as inappropriate continuation despite a weak possibility of showing a benefit from treatment**

In many clinical trials it is useful to perform an interim analysis on the data to assess the probability of obtaining a statistically significant result if the study is continued to the end. If this probability is very low, one may decide to save resources by stopping the study before it is finished even though it has not provided a convincing answer. It is in this sense that ‘futility analyses’ are usually understood. However this approach may not be appropriate in this particular study for a number of reasons:

- this concept of termination for futility is appropriate for trials concerning two competing (one innovative, the other standard) treatments that are approximately equally effective, with outcomes that occur early and outcome rates that remain relatively constant over time, and where the study would almost certainly continue to the maximal sample size were the null hypothesis found correct at interim analyses [10]. These assumptions do not hold in this trial. In the present study adverse events in the ‘treatment’ group will occur immediately at the time of intervention, during the recruitment years, while hemorrhagic events in the ‘deferred treatment’ group will occur throughout the study. We expect the rate of events in the deferred treatment group to be low (between 0.5 and 2% per year) and hemorrhages to occur at random, without any logical relation with time since diagnosis, but this remains unknown (another, more mechanistic, wear and tear model could hypothesize that the parietal weakness leads to increasing risks with time). By chance alone, there could be years of low and years of high rates of hemorrhage, especially when n is small at the beginning of the trial (for example for 200 patients in each group). The risks of treatment, immediately at the time of intervention, are meant to prevent hemorrhages during the following years, but probably correspond in magnitude to years of hemorrhagic risks. Patients allocated to the ‘treatment group’ may present a recurrence, and the rate of recurrence increases with time [5]. Thus we cannot assume that hazard ratios remain stable or that a benefit of treatment noticed at a point relatively early in the trial can be projected for the remainder of the observation period. In the treated group, the confidence interval for observed rates of severe complications would still be compatible with the primary hypothesis are very wide when n is small. For example after recruitment of 400 patients and one year of follow-up for all, the number of events in the conservative group is expected to be between 0 and 8, the number of complications in the treated group between 4 and 24 (exact binomial tests). With so small numbers, it is easy to conceive that by chance grouping of events alone certain years could appear favourable or harmful for one group or the other and interpretation of early statistics should be very conservative;
- a fundamental assumption of any futility analysis procedure is that the population after the interim analysis is expected to be similar to one preceding the analysis [4]. In some studies subjects enrolled early may be to some extent different from subjects enrolled later. This may very well be the case in this trial, in which we expect an initial reluctance to question received opinions and a significant discomfort with randomization for many clinicians in the field. From our own experience we expect the ‘comfort zone’ regarding the two options to increase with time as clinicians free themselves from their intuitive clinical decision-making habits [7]. We strongly encourage investigators to consider enrolment for any patient up to now considered for treatment, but we expect an initial reluctance, with a preference for lesions with an a priori lesser risk of bleeding (small anterior circulation aneurysms as compared to large posterior circulation aneurysms for example). This type of ‘cherry picking’ is discouraged, but because we need the participation of 60 centres of various cultural and demographic backgrounds, the ‘type’ of patients that will be recruited early on remains unpredictable. Hence we suspect the subjects will significantly differ as we progress and the assumption that interim results early on can be projected in the future is likely to be false;
- this concept of futility is particularly relevant when the distinction between regimens that are as effective as standard care and those that are inferior has no importance, and the trial may not be continued to make this distinction. However, ‘this distinction between ineffective and harmful therapy becomes important when the therapy is already in wide use and thought to be effective by many physicians and leaders of the field’ [10]. In this research effort, facing the received opinion that treatment is helpful, it may be equally important to pursue our effort until we get convincing evidence that treatment is harmful, and not leave the subject matter with an equivocal answer. Interruption for futility is currently a controversial issue but even those authors favouring futility stopping emphasize that ‘Futility-stopping rules should not be used when large segments of the community already believe that a treatment is effective [8];
- another inappropriate use of this concept of futility is when factors other than success or failure of the treatment under study should affect decisions regarding termination: ‘a common situation is that in which the data collected in a randomized clinical trial provide valuable information about the natural history of a disease [10]. This point is particularly relevant here, since there is no valid data regarding the natural history of unruptured aneurysms, apart from non-randomized registries;
- finally other reasons for being reluctant to prematurely interrupt such a trial include the potential direct clinical impact of the trial on treatment policies and the very low likelihood of repeating such a trial [1]. The secondary objectives of the trial, including knowledge of the natural history of the disease, safety and efficacy of endovascular treatment, morphological evolution of
lesions at imaging, neuropsychological and psychological impacts of knowing to have an intracranial aneurysm and quality of life data are most valuable knowledge that would lost if the trial was interrupted on the basis of futility criteria restricted to the primary hypothesis.

Hence we conclude from this section that interruption because of futility, understood as a weak chance of showing a benefit of treatment during a trial progressing normally, is unlikely to be appropriate.

It the DSMC requires this type of futility analysis, there are many possible approaches: Stochastic curtailment, based on conditional probability of rejecting the null hypothesis (the problem with the approach is that the difference between the 2 treatments need to be specified in advance), and the Predictive Probability of Significance, the weighted average of conditional probability of rejecting H0. Both are group sequential tests that are not convenient if the response time relative to the accrual rate is not short, which is the case in this particular trial. We propose to verify that observed frequencies are within the confidence intervals of the ones used in the calculation of the sample size necessary to reach the primary objective. Sample size adjustments could be performed once a significant number of subjects have been followed for at least one year. We also propose to limit the comparison between the 2 groups to 2 interim analyses to be performed once recruitment is finished and to use the very conservative Bonferroni method to correct for multiplicity.

**Interruption because treatment is harmful**

Endovascular treatment of unruptured aneurysms is considered in most centres ‘standard practice’. The DSMC is not facing the situation where it has to sanction the use of an innovation. The trial was designed to compare elective coil- ing and conservative management, which may include treating aneurysms in the same fashion but once clinical indications are thought to occur. Although it is unlikely that indications will occur in a majority of patients, there is much uncertainty regarding the frequency of this occurrence. This creates tension regarding the responsibility of stopping the trial because the treatment is judged harmful (because of complications that are judged excessive as compared to benefits or relative to the risks of observation) and the fact that a certain proportion of observed patients will end-up necessitating to the same ‘harmful’ treatment after a period of conservative management.

Another way of looking at this problem is to estimate what kind of information would arbitrarily be felt convincing, or clinically meaningful, such as finding that the rate of permanent deficit from treatment is much higher than previously thought, or as high or higher than other approaches, perhaps more effective, such as surgical clipping.

**Interruption because treatment is harmful**

Interruption because treatment is judged to be harmful, after 15 years of widespread clinical use by the expert community, necessitates very convincing evidence. In descending order of force, the following are thought to be inescapable conclusions: i) The rate of hemorrhagic events is found to be as high or higher after treatment than the one found in the conservative group. In other words, even if there were 0 complications, there would be no protection from hemorrhages. ii) Treatment is judged harmful because initial significant complication rates exceed the rate of hemorrhages in the patients observed for 10 years, while efficacy remains under 100%.

These arguments in turn depend on the observed frequency of hemorrhagic events in the conservative group. Because frequencies are expected to be low (between 0.5 and 2%/ year), this rate can only be validated after a substantial numbers of subjects have been observed for a meaningful period, likely to extend beyond the recruitment period, when no further subject will be included into the trial. Hence stopping rules should in this particular trial be set at a conservative level, and stopping for safety reasons is very unlikely.

**Interruption because hemorrhagic events are too frequent or too infrequent in the conservative group**

Most errors in estimating frequencies in the planning of clinical trials are in the direction of overestimating the frequency of the outcomes. Given the less likely possibility that we have underestimated the risks of observing patients with unruptured aneurysms, it is unclear how one is to judge the frequency of ruptures to be unacceptably high without knowing the risks and efficacy of treatment, or before actually going through the trial itself.

Could there be such a number of ruptures that the DSMC would judge that, no matter what is the efficacy or the safety of treatment, coiling should be considered mandatory and the trial should be stopped? Given this reflection and the fact that treatment is commonly thought to be standard of care only emphasizes how badly needed this trial is. Once again we propose that stopping rules for this possibility should be extremely conservative, at least during the recruitment phase of the trial. Then this question can be conflated with the one addressed in the next section: Interruption because treatments are convincingly different.

Conversely one word must be said about the possibility for events to be insufficiently frequent for the trial to be meaningful. One of the reasons to do the trial is to assure that the estimated risks of the disease concern patients that are actually considered for treatment. The ISUIA registry is felt to be unbelievable by many because the lack of randomization makes one suspect that patients that were included in the conservative arm were those in whom treatment was not felt to be indicated, thus introducing a strong bias for lesions at low risk of hemorrhage. Still our primary hypothesis is based on incidences reported for the ISUIA ‘conservative’ group. Thus it is unlikely that number of events would be lower in this randomized trial where investigators can only include patients in whom they are committed to coiling should randomization indicate so. Annual estimates of the hemorrhagic risk are already low, so much so that confidence intervals for observed rates when the sample is small become almost meaningless (after 400 patients the number of events in the conservative group (200) can be anywhere from 0 to 8). The claim that the
rate of ruptures is so low that recruitment should be stopped, is unlikely to be asserted with confidence using a fraction of the total sample observed during the small number of years of recruitment (as compared to the 10 years of observation).

One possibility is that investigators, reluctant to question the soundness of treating unruptured aneurysms, systematically recruit only patients in whom risks are felt to be very low. One way to control this potential circumstance is to verify the demographic data, looking for abnormal frequencies of suspected risk factors as compared to standard endovascular series.

**Interruption because treatments are convincingly different**

The primary hypothesis of the trial is that it will show at its completion that patients treated with coiling have a better outcome, but the efficacy of endovascular coiling in the prevention of ruptures has never been established. Because patients of the treated group are initially submitted to the risks of treatment, we anticipate an initial difference in favour of conservative management that could only be compensated with time (Fig. 1). In addition, we expect a certain proportion of angiographic recurrences, between 10-20%; increasing with time after coiling; presumably these patients may no longer be protected from ruptures, but these clinical events can only be captured in a delayed fashion as compared to events in the ‘conservative’ group, where risks of rupture presumably remain stable throughout the duration of the follow-up period. A certain proportion of patients in the ‘conservative’ group will eventually be treated. Hence risks of treatment may also occur in the so-called ‘conservative’ group, but in a delayed fashion as compared to the endovascular group. For these patients one may wonder if risks are going to be taken anyway, why not initially to benefit from protection from hemorrhage earlier?

For all these reasons, the essential assumption for early interruption of trials, that is ‘if the hazard ratios remain the same’ and benefits can be projected in the future for unobserved subjects or unobserved time periods, is unlikely to be true. Thus by design, early interruption because the conservative group fares better is not permissible, and early interruption because the endovascular group fares better is unlikely early on, and should be done in a very conservative fashion later on. Early interruption because the endovascular group fares better is equivalent to recommending treatment for all patients, and presumably this should not be proposed before ensuring that the risks of hemorrhages from recurrences are not excessive and negating earlier benefits.

Hence we believe that sequential, group-sequential, alpha-spending functions, techniques based on conditional probabilities and Bayesian approaches are inappropriate for this trial. We rather propose a very classical, conservative frequentist approach. Interim monitoring during the enrolment phase should be limited to assuring that observed frequencies are within confidence intervals of the study hypotheses. At some point before the conclusion of the recruitment period, observed frequencies should be used to recalculate if the sample size is compatible with a meaningful trial. Comparative looks should be limited to the follow-up phase, and taking care to adjust for the risks of type I errors. We propose a total of 2 interim looks and one final analysis for the primary outcome. We calculated the sample size to correct for multiplicity using the very conservative Bonferroni method, requiring us to restrict the two-sided alpha to 0.0167 for each analysis. Of course, because the significant action that can be done by the DSMC during the follow-up phase is recommending that patients of the ‘conservative group’ be offered treatment, only results convincingly supporting active treatment will be meaningful. By the nature of the problem and the design of the trial, a public warning that endovascular treatment may be convincingly demonstrated as beneficial, or harmful, will likely have to wait for the final statistical analyses.

**Concluding remarks**

Anticipation of potential results of a large clinical trial is an important exercise that permits to foresee difficulties and helps to better understand what will be the meaning of the potential outcomes. From the previous discussion, we conclude that futility because of insufficient recruitment is the single realistic threat to the realization of this trial.

Recruitment may be jeopardized by fear of the disease, faith in our technologies, and discomfort with the difficult notion of equipoise and uncertainty. A radical modification of mentalities is needed to make this enterprise a success.

Once recruitment proceeds satisfactorily, trial interruption is unlikely. As long as we remember that the DSMC is not sanctioning the potential dangerous use of an innovative treatment, and that all along monitoring assures that
the treated group is not doing better than the conservative group, the study of randomized cohorts of treated and untreated patients would bring so valuable and so unique a knowledge that we believe the trial should be pursued even though interim analysis suggest that the probabilities of reaching a convincing answer become low.

Some say we are treating populations, while they wish to treat individuals. But unless they possess some power of divination, nothing about the future of the individual can be known in advance. Wishful thinking is a poor justification for risky treatments. The trial has been designed to optimize the outcome of each individual participant, and monitoring will assure the well-being of all participants. Patients will be recruited one by one and aneurysms will be treated one by one. The best we can do is to proceed with the study and hope to show that one treatment, in general, leads to a better outcome.

Patients with unruptured intracranial aneurysms are healthy individuals. What we are treating, when treatment is elected, is a potential risk. The magnitude of this risk must be evaluated in a rational fashion, and balanced against the risks and benefits of treatment. No matter how reluctant we are to resort to statistics, they are inescapable. Some statistics from registries already exist; resorting to these statistics does not require the hard work and the change in mentalities involved in organizing and implementing a randomized trial. Either we trust them, and renounce to offer preventive treatment (because registries have shown that treatment is rarely indicated) or we hypothesize that data from registries is unreliable, and we do what is necessary to have reliable evidence to guide our actions.

Acknowledgments

The study is supported by a grant from the Canadian Institutes of Health Research (MCT- 80799).

International Standard Randomised Controlled Trial Number (ISRCTN): ISRCTN62758344.

References


[8] Schoenfeld DA. Pro/con clinical debate: it is acceptable to stop large multicentre randomized controlled trials at interim analysis for futility. Pro: futility stopping can speed up the development of effective treatments. Crit Care 2005;9:34-6 (discussion 34-36).

