REVIEW

Unruptured intracranial aneurysms: Their illusive natural history and why subgroup statistics cannot provide normative criteria for clinical decisions or selection criteria for a randomized trial

Anévrismes intracrâniens non rompus : leur histoire naturelle trompeuse et pourquoi les statistiques de sous-groupes ne peuvent fournir des critères dans les normes pour les décisions cliniques ou pour les critères de sélection dans un essai randomisé

J. Raymond\textsuperscript{a,∗}, D. Roy\textsuperscript{a}, A. Weill\textsuperscript{a}, F. Guilbert\textsuperscript{a}, T. Nguyen\textsuperscript{a}, A.J. Molyneux\textsuperscript{b}, A.J. Fox\textsuperscript{c}, S.C. Johnston\textsuperscript{d}, C. Cognard\textsuperscript{e}, L. Pierot\textsuperscript{f}, J.-F. Meder\textsuperscript{g}, Isabelle Rouleau\textsuperscript{h,i},

Trial on Endovascular Aneurysm Management (TEAM) collaborative group

\textsuperscript{a} Interventional Neuroradiology Research Unit, Department of Radiology, université de Montréal, CHUM Notre-Dame hospital, 1560, Sherbrooke east, suite Z12909, H2L 4M1 Montréal, Quebec, Canada
\textsuperscript{b} Neurovascular Research Unit, Nuffield Department of Surgery, University of Oxford, Radcliffe Infirmary, OX2 6HA Oxford, United Kingdom
\textsuperscript{c} Department of Medical Imaging, Sunnybrook and Women’s College Health Sciences Centre, M4N 3M5 Ontario, Canada
\textsuperscript{d} Department of Neurology, University of California, 94143-0114 San Francisco, California, USA
\textsuperscript{e} Neuroradiologie diagnostique et thérapeutique, hôtel Purpan, CHU de Toulouse, 31059 Toulouse, France
\textsuperscript{f} Service de radiologie, hôpital Maison-Blanche, CHU de Reims, 51092 Reims, France
\textsuperscript{g} Département d’imagerie morphologique et fonctionnelle, centre hospitalier Sainte-Anne, 75014 Paris, France
\textsuperscript{h} Département de psychologie, centre de neurosciences de la cognition, université du Québec, H2X 1L7 Montréal, Quebec, Canada
\textsuperscript{i} Service de neuropsychologie, hôpital Notre-Dame, CHUM, H2L 4M1 Montréal, Quebec, Canada

Available online 1 February 2008

∗ Corresponding author.
E-mail address: dr_jean_raymond@hotmail.com (J. Raymond).
URL: http://www.TEAMstudy.org (J. Raymond).

0150-9861/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.neurad.2007.12.001
Very little is known about unruptured intracranial aneurysms (UIA). As their name implies, they seem to be predestined to rupture, but the frequency of this dreadful event and the appropriate management of the individuals carrying aneurysms remain controversial.

Experts in the treatment of aneurysms have to deal with the consequences of ruptures on a daily basis. Despite decades of progress in the clinical management of these patients, an honest evaluation of the clinical outcome after aneurismal rupture is a sobering experience. Treating UIA is so gratifying as compared to the management of ruptured aneurysms that we must remind ourselves that such comparisons are invalid, and refrain from the temptations of preventive actions on this basis alone. Furthermore, treatment carries definite risks of disability and mortality. Hence, it is yet impossible to know today whether, by treating UIA pre emptively, we do harm or good.

One alleged solution to the dilemma is to study the so-called natural history of the disease, and to compare the incidence of hemorrhage in untreated patients with the risks of treatment in other patients [1]. This invalid procedure could still perhaps provide a supportive argument, but in one direction only: “if the magnitude of the difference between the best available natural history data concerning these patients and the morbidity and mortality associated with repair was such that one would be taking more than a normal lifetime of risk in one day to surgically repair such a lesion” [2], one could propose that conservative management offers the best overall option. But a recommendation in favor of a risky treatment performed in a context of prevention would necessitate much stronger evidential proof of benefit, such as can only be provided by a randomized trial, as required by modern medicine [3,4].

There are at least two problems with this approach:

- if the demonstration that a new therapy carries more risk than the "natural history" of a disease, this may temper or discourage its adoption by the expert community, much more convincing arguments are needed to condemn a treatment that has become entrenched after more than a decade of use and refinements;
- more importantly, the so-called "natural history" of unruptured aneurysms is at best a confused notion and we will contend an illusive concept. In line with the pre-scientific notion of natural history, observational studies have attempted to classify aneurysms into well-defined...
categories or "natural kinds" purported to carry a high or low hemorrhagic risk. We will review reasons not to rely on such categories.

Another goal of this manuscript is to review the kind of evidence available on subgroups of aneurysms. We will assert that available statistics issued from doubtful methodology can neither serve as a normative basis for clinical decisions nor give boundaries that could help choose appropriate selection criteria of a valid trial.

The natural history of the disease
We assume that the natural history of a disease is commonly and quite clearly understood by most physicians as "what usually happens if no action is taken". In this way, natural refers to a usual (common/most frequent) evolution that occurs when things remain undisturbed by artificial or human interventions. Things, however, become muddy as we try to look into the concept with any precision. One key word here is "usual", but the gist of our argument will rather be revolving on the "common/frequent" complex, combining a notion of repetition of an outcome (frequent) with the one of belonging to a group of individuals sharing certain features (common).

Natural history is an old concept, which goes back to Aristotle and other ancient philosophers, owed and perplexed by the diversity of the living world. This overwhelming diversity prompted many attempts at classifying species into taxonomic groups, culminating in the system of Linnaeus in the 18th century. There are many problems in such analysis by observation, constructing groups by resemblance and separating individuals according to certain, often arbitrary differences. Furthermore, all classifications presuppose an order, natural, supernatural, or more likely conceptual (man-made). Early taxonomists acknowledged that if "man is forced to use particular methods, to arrange in an orderly way the infinitely numerous and varied objects which he examines, to make clear distinctions among the immense multitude of these things", "nature has never made anything like this, and instead of doing ourselves an injustice and confusing our works with hers, we must recognize that in this business the classes, orders, families, genera, and nomenclatures are inventions of ours" [5].

Hence, we must be careful in attributing characteristics to the nature of things and attempt to distinguish what could be artificially constructed by our man-made methods.

Assuming that aneurysms constitute a defined "natural kind", what could then be the natural history of aneurysms? Perhaps this concept could refer to knowledge of the causal factors responsible for the development of dilatations of arteries, when and how they form, and what usually precipitates their rupture. But there is no guarantee that aneurysms behave like species, share a common history, observable in a normal environment, according to some natural law. For all we know, aneurysms may be accidents that simply behave like stochastic phenomena, rupturing at random, evading, forever, any attempts at reliable individual predictions.

A scientific determination of what could be the natural history of diseases is fraught with multiple difficulties: the choice of models, the identification of covariates, the nonnegligible errors intrinsic to screening tests, the mathematical and time sequence assumptions, the concomitant therapeutic interventions, to mention a few, will significantly affect results [6].

Things become more complicated and confused if we stretch the concept, advance by one more step and pretend that there are multiple "kinds" of aneurysms. Even if we limit our inquiry to berry aneurysms unassociated with congenital disorders or other known causal factors, there is a wide variety of lesions differing in morphology, multiplicity, location and size. These differences do not necessarily translate into distinctive histories; most lesions remain silent until they rupture. Hence, natural history in this field solely refers to a risk of rupture, a frequency of a dreaded event, defined as a ratio of a number of ruptures, on a denominator, the number of patients identified with the disease, observed over a certain period of time.

How natural could the natural history of an imaging finding be? Presuming that the event we wish to prevent could not have gone unnoticed (ruptures, deaths or significant morbidity), the fact that noninvasive neurovascular imaging, increasingly accurate and available over the last decades, has revealed an ever increasing number of unruptured aneurysms in patients referred for unrelated symptoms, can only play in favor of a benign natural history, the denominator rapidly growing in the presence of a stable numerator. This is exactly the perception of the medical community, but note that this ratio reveals more about the evolution of medical imaging, referral patterns for imaging, and comorbidities, than it discloses the hypothetical essence of "natural kinds" of aneurysms. If there is such a thing as a natural history of aneurysms, it cannot be defined by such a ratio, which depends on the construction and clinical use of sophisticated machines. In addition, far from being natural, this ratio will evidently depend on the research methodology used to provide those numbers. For example, the results of the retrospective phase of the ISUIA study (International study on unruptured intracranial aneurysms), published in 1998 [1] were greeted with appropriate scepticism and a flock of indignant editorials. This retrospective study reported a low (0.05% per year) risk of rupture for the most frequent lesions, the small (then defined as < 10 mm) anterior circulation aneurysm in patients without a history of rupture of another lesion. This study allowed recruitment of patients with prevalent aneurysms, and the posterior recording of years of event-free intervals the years the patient had been known to carry an aneurysm. Hence, this extraordinary low frequency probably reflects the simple fact that other patients who may have died from SAH during the past years cannot present at the clinic today for unrelated problems.

The prospective phase of ISUIA was more interesting [7]. Patients selected for observation, clipping or coiling were followed for a mean of four years. It is presumed that patients that were observed can provide us with data regarding "natural history" of the disease, but in fact this is solely a name given to the evolution of an undefined group of patients that were excluded from treatment. Far from being undisturbed by human interventions, the resulting data will evidently depend on who, when and for what reasons patients were excluded to be attributed to the other two groups. The overall incidence of rupture in the prospec-
tive arm of ISUIA was in line with previous publications: at least 51 events were recorded in 6544 patient-years, for a yearly incidence of 0.8% (0.6–1.0%).

Different “kinds” of aneurysms

If ISUIA provided similar data to the previous publications, why did it stir so much controversy? The main reason, we contend, is because the publication attempted to draw firm conclusions about the management of small aneurysms. However, every week experts in the treatment of aneurysms are confronted with patients that suffer devastating hemorrhages from small aneurysms. In addition, some published series on unruptured aneurysms followed for a long period of time have shown that the majority of ruptures occurred in aneurysms of 6 mm and less, in size [8].

It is true, ISUIA showed a very low incidence of events in certain categories of patients. But a critical reading of the data could go as such: if you observe patients selected for observation, in poor health (five-year mortality 12.7%), with a disproportion of smaller less symptomatic lesions; if you include extradural lesions, but exclude from analyses many intracranial hemorrhages that you observed (it is unclear how many according to the text, but at least 31), project results for five years even though 33% of patients were eventually operated on and less than 20% followed for four years; if you divide the small number of selected events into at least 15 data-driven categories according to size and location; if you systematically attribute the hemorrhage to the larger lesion or to posterior circulation lesions when patients have multiple lesions (40% of patients); and finally if you arbitrarily exclude from one category lesions that happened to be prevalent and the ones associated with the events (Pcom aneurysms excluded from anterior circulation aneurysms), you can show that the resulting risk of rupture for remaining patients is very small, or close to zero. This procedure is like claiming that the rupture risk of a group of patients is close to zero when you exclude patients that ruptured. In addition, the data is presented in such a manner as to prevent the objective evaluation of the reader, by omitting to disclose absolute real numbers (certainly very small in many columns of Table 4 [7]), while not reporting other meaningful information, such as confidence intervals that certainly would have shown overlaps between all artificially created categories. For example, a recent meta-analysis of the literature [9] reports yearly risks of 0.5 (0.25–1.0%), 1.2 (0.65–2.0%) and 1.5% (1.1–2.0%) for less than 5 mm, 5–10 mm, and greater than 10 mm aneurysms, respectively (C.I. calculated from Table 3). On must remember that if one wishes to convincingly claim that one group carries a 1% risk, another 0.5%, more than 4670 patients per group must be recruited (using 80% power and 5% alpha), a feat yet to be accomplished in this field.

It is sad to realize that the ISUIA report on subgroups of aneurysms does not pass muster standards of clinical research [10]. Available observational studies do not have the power, the accuracy or the appropriate methodology to provide the “truth” about small aneurysms, or a risk that could be reliably projected to future patients.

Going back to the notion that a natural history study could give us “natural kinds” of aneurysms, as defined by significantly different essences, we must attempt to sort out what is out there in the world from what is constructed by our methods. One must be careful in anticipating the effects of a classification of patients or risk factors on data itself, and this has not been sufficiently stressed before. The outcome concerns patients, but patients were classified according to aneurysm characteristics. Hence, a patient could be a member of a number of different classes. Forty percent of patients observed in ISUIA had multiple aneurysms; although no information is given in the article on the treatment of this problem, one can speculate from Table 4, in which the total number of group members matches the number of participants, that patients with more than one lesion were categorized according to a “dominant aneurysm”. Now, how this dominance is determined will artificially influence results. If by convention a patient with both a small and a large aneurysm is categorized as a patient with a large aneurysm, or one with both an anterior and a posterior circulation aneurysm as a patient with a posterior circulation aneurysm, then any hemorrhagic event that could occur will automatically be attributed to the “dominant” aneurysm. Hence, even before the trial starts, one will expect a priori a significant difference between small and large aneurysms, and between anterior and posterior circulation aneurysms, because this will be true “by definition”. This is an effect of the classification scheme, not a valid observational truth concerning aneurysms. Of course, another net effect will be an artificially low relative risk for small anterior circulation aneurysms. If you are allowed to exclude Pcom aneurysms from carotid aneurysms, what happens to posterior circulation aneurysms if you exclude basilar bifurcation aneurysms? Do they still have a higher risk? There are at least 20 typical locations of aneurysms with a varying prevalence; how can we deal with this?

Everyone agrees that in order to apply a certain statistical information to a particular patient, one needs pertinent statistics regarding the narrowest reference class possible in which the patient is a member, but we rarely have the luxury of possessing valid statistics for multiple reference classes and their combinations. This would necessitate impossibly large trials.

The principle of a “natural history” methodology is to attempt to correlate some perceptible characteristic (morphology, location, size, etc.) with some crucial event that occurred in observed cases and that is of interest to us (hemorrhages). Most importantly, the correlation must be reliable and repeatable enough to allow induction or prediction of future meaningful events in unobserved cases. For example, certain kinds of edible mushrooms can be differentiated morphologically from poisonous ones. With modern biochemistry, the conditional “if this is eaten, a sickness ensues” can now be explained by the invisible presence of a toxic molecule in certain kinds of mushrooms. However, we doubt that there are kinds of aneurysms like there are kinds of mushrooms. And we doubt that statements like “some aneurysms can rupture (risks 1–2% per year) while others do not (risks 0.25–1.0%)” will ever be as meaningful or as true as statements like “some mushrooms are edible and others are poisonous”.

We believe that if aneurysms are a “natural kind”, their essence is to present a risk of rupture. Now risks mean ratios of collected cases, and there is nothing natural in our methods to separate or lump collection of cases.
We seem to have ample evidence that the risk of rupture increases with the size of aneurysms, but we should not succumb to the temptation to search for precision where only vagueness can be found. Our classifications may, in the words of Lamarck (1809), "constitute the artistic parts in the natural sciences, parts which one must be careful not to confuse with the laws and the acts of nature herself" [5]. This warning was also stressed by Buffon (1749) who insisted that "our systems are built on uncertain ground... and can only show men’s propensity to find resemblance between most different objects and regularity where only variety can be found" [11].

Designing a study and estimating sample sizes necessary to answer not only the main research question, but other interesting subgroup findings as well, is rapidly a sobering experience. The only practical option is to design trials to answer the most important and generalizable question, and to verify if there are subgroups for which results are vastly different from the main finding. Subgroup membership must be determined before initiation of the trial and prospective allocation of type I and II error rates is needed if the subgroup results are meant to have confirmatory value [12]. If there are important differences that are discovered post-hoc, one should be very careful in interpreting the data, because the analysis is solely exploratory, and subject to sampling error. In any case, "the best descriptor of a subgroup is the finding observed in the overall cohort" [13]. At most, the subgroup of interest could be the object of another trial. But this data should never serve to invalidate the main finding. And post-hoc subgroup findings of observational studies cannot provide a valid justification for risky clinical decisions. The sampling errors and biases, inevitably included in the subgroup statistics that are currently available, will be multiplied in an uncontrollable fashion when inserted in our homemade algorithms designed to estimate life-time risks. These mathematical chimeras will then be compared with immediate treatment risks (also determined by dissections of case series vastly different from natural history series), to provide a pseudoscientific calculus supposedly tailored to the individual, the current basis of our unwarranted expert opinions and illusory certainties.

Starting in the 19th century, Natural History, formally taught by college professors, was progressively replaced by more specialized sciences and relegated to an amateur activity rather than a part of science proper. Our quest for an evanescent natural history of kinds of aneurysms was ill-advised, and the concept should find its place in natural history museums. The clinical problem calls for a different approach. If what we need is evidence to justify clinical decisions, a direct comparison between two treatments allocated at random is the standard method of modern medicine.

Why subgroup results cannot provide useful inclusion criteria of a randomized trial?

If exploratory findings cannot provide normative criteria for clinical decisions, can we at least use them to help design the trial necessary to solve the clinical dilemma regarding the treatment of unruptured aneurysms? There are many arguments for not opting for aneurysm size categories as selection criteria for a trial, in addition to their unreliability mentioned above.

The pragmatic argument

Let’s imagine we know the result of a 10–15 year trial, in which a size boundary was arbitrarily chosen as an inclusion criterion (for example, all aneurysms < 7 mm were excluded). Let’s say the results are that patients with 7 mm aneurysms have a large benefit from treatment. What about 6.5 mm aneurysms? Or the converse scenario: we excluded aneurysms greater than 7 mm because treatment was deemed necessary; we show at the end of the trial that aneurysms less than 7 mm should not be treated. Then, what about 7.5 mm aneurysms that were always treated without evidence? Should we initiate a new 10–15 year trial?

The 7 mm boundary is exactly at the mean of all lesions observed in ISUIA. The median size of treated unruptured aneurysms is 7 or 8 mm [14]; can we exclude 50% of cases even before initiation of a scientific study, and if we do, on what basis? If scientific evidence is supposed to provide the reliable evidence, the normative criteria cannot come before the trial!

The research question should address the most common patients, and provide a generalizable answer to the most prevalent problem.

The argument of the imaging specialist

Aneurysm measurements can significantly vary according to imaging modality, equipments, observer-related factors, projections, etc... Given all these causes of variations, we cannot expect a precision beyond the millimetric range. Assuming that the risks are distributed in a continuous fashion according to size, it seems risky to dictate a precise size criterion right at the apex of a Gaussian distribution. A size criterion would seem to make sense only at the extreme of a normal distribution, in order to limit arbitrary exclusions to the minimal number of subjects for whom the research question is pertinent.

The soritic argument

In close relation with the argument of the imaging specialist, is the one inspired from the paradox of the Sorites (how many grains constitute a heap, or how much hair must be lost to qualify as bald?), an ancient puzzle about vague predicates that resist setting a fixed boundary. We intuitively believe that if risks are related to size (for example), this relation must hold along a continuous function, from a low risk in small aneurysms to a high risk in large aneurysms, but with no precise threshold. Despite attempts at solving the paradox using fuzzy logic or mereology, in such a context most people are comfortable applying such predicates only to the extremes of the scale.

The argument of the pertinence of other factors

Most clinicians would consider many patient and aneurysm-related factors in their clinical decision: lesion size,
Unruptured intracranial aneurysm. Their illusive natural history

location, previous history, concomitant diseases, treatment risks, age and life expectancy, etc. Other factors have been incriminated in the past: midline location (perhaps a misconceived criterion based on the combination (ACom + basilar bifurcation) rupture more frequently and MCA aneurysms bleed less frequently), the presence of daughter sacs, smoking, heavy drinking, etc. If there are other factors that should be taken into our calculus, then size cannot be normative on its own; for example, if treatment is warranted in 60-year-old patients with 7 mm aneurysms, then perhaps treatment is warranted for 50-year-old patients with 6 mm aneurysms, or 30-year-old with 5 mm aneurysms.

Then, if the 7 mm boundary is taken as ‘‘truth’’, what about location or history of hemorrhage? If a 8 mm MCA aneurysm should be included, what about a 6 mm basilar bifurcation? Then, perhaps we would need a complex set of inclusion criteria with different sizes for different locations, and different size/location combinations for patients with different history and a different algorithm for patients at various age... However, large (>7 mm), group 2 aneurysms (anterior or posterior location) bleed less frequently (2/115) than small (<7 mm), group 1 posterior circulation aneurysms in the ISUIA registry [1]. Should this finding be included in our algorithm? Now we are formally building bias from previous registries in the design of our trial that was supposed to minimize bias! Can the trial designer, arbitrarily and authoritatively decide before the trial starts the appropriate combination of subgroups and a precise algorithm that could skew the data? This would clearly be antiscientific.

The argument from multiplicity

If the risk of a greater than 7 mm aneurysm is too small to consider inclusion into the trial, what about a patient with two aneurysms each 6 mm in size? As many as 40% of patients have multiple lesions [1]. How should we apply our imaginary algorithm in the case of multiple lesions?

The argument of the DSMC

If knowledge of subgroups is crucial, then when is the DSMC supposed to stop a trial that has shown a large benefit from treatment? If a large benefit is shown for the general population, is it shown for lesions less than 7 mm as compared to those that are greater than 7 mm? For this or that subgroup? No, then let us continue the trial statistical evidence for every potentially pertinent category (size, location, previous history, morphology, sex, age...). But each time we split the general data into categories, we get data that is no longer statistically significant. This is alike the Heisenberg’s principle of uncertainty applied to probability: as we narrow down the reference class, sampling error increases and we lose precision. Should we continue the trial for ever, looking for more and more precise subcategories of patients, not allowing any generalization, and in the meantime submitting patients to suboptimal treatment despite the scientific evidence proven for the majority?

What should then be the optimal inclusion criteria of a valid trial?

If we keep in mind our main purpose, to solve the clinical dilemma regarding the treatment of unruptured aneurysms, we shall opt for a ‘‘management’’ type of trial. This call for a large, simple trial, looking for a pragmatic answer, with loose eligibility criteria based on uncertainty, taking all comers, retaining every admitted patient in the analysis, ascertaining a range of hard outcome events and counting every event and charging it against intervention [15].

No doubt participating investigators will somewhere in their individual calculus take into account the published findings from available observational studies to decide if the appropriate treatment of an individual subject with an unruptured aneurysm is currently uncertain. We must, however, accept the limits of what can be known in a realistic world. We are looking for a general answer to the best approach to unruptured aneurysms, and may never reach precise and definite answers for all interesting categories of subjects.

If it is unwise to rigidly apply subgroup findings from observational studies to selection criteria of a trial, is it wise to use them in guiding our clinical decisions, before any valid demonstration of the benefit or harm caused by preventive treatment of aneurysms?

Conclusion

Our previous endeavours to capture a ‘‘natural history’’ of unruptured aneurysms were ill-advised. Numbers on unruptured aneurysms are man-made, and depend on our routine use of technology and clinical decisions. Available statistics on subgroups of unruptured aneurysms are unreliable and cannot provide a prudent and rational basis for clinical decisions or reasons to arbitrarily exclude large segments of the population for whom the research question of a randomized trial would be pertinent.

Acknowledgements

The study is supported by a grant from the Canadian Institutes of Health Research (MCT- 80799) and registered International Standard Randomised Controlled Trial Number (ISRCTN): ISRCTN62758344.

References

and reasons to participate in a randomized trial. J Neuroradiol 2006;33:211–9.


[5] Lamarck JB. Zoological Philosophy. Part I; Chapter One: On the Role of Art in the Productions of Nature. Source: http://www.mala.bc.ca/~johnto/lamarck/tofc.htm (accessed on June 18th, 2007). [This translation has been prepared by Ian Johnston, of Malaspina University-College, Nanaimo, BC, Canada. All sections listed below are in the public domain, and may be used by anyone, in whole or in part, without permission and without charge, provided the source is acknowledged, released September 1999.].


