Small Aortic Aneurysms: Is Evidence Evident?

A. Branchereau

Faculté de Médecine de Marseille, Université de la Méditerranée, Assistance Publique Hôpitaux de Marseille – Hôpital La Timone, Département de Chirurgie Vasculaire, Marseille, France

Key Words: Aortic aneurysm; Small aneurysm; Surveillance.

The results of the UK small aneurysm trial were published for the first time in 1998,1 and the long-term results were published in 2002.2 The crude results presented in the above two articles may be summarized as follows. No significant difference in overall survival between the two groups;1 impaired quality of life in the surveillance group;3 improved long-term survival by early surgery.2 The investigators conclude: ‘results do not support a policy of open surgical repair for abdominal aortic aneurysms of 4.0–5.5 cm in diameter’.1

One critical remark made immediately was the 5.8% mortality observed after elective surgery which is higher than usual in major vascular surgical centres; Nevertheless, we agree with the investigators that, if this rate is higher than the ones observed in some reference centres,4 it reflects reality in the geographical area where the trial is conducted and is comparable with the mortality rate observed in community studies.5 Moreover, the results of the ADAM US study following the same design were published in 2002. With a 2.7% in-hospital surgical mortality no difference was observed between the two groups.6

Concluding, evidence-based medicine cut it short: no benefit from surgery versus surveillance for aneurysms below 55 mm in diameter. Are things that simple?

External Validity of the Design

The trial was designed to have 80% power to show a significant difference at the 5% level.1 The investigators concluded that there was no difference between the two groups. They proceeded like they would have in a trial of equivalence although they deal with a trial conceived to show superiority. Question: is there really no difference or is the power of the study insufficient to show a difference?

In 2002, the long-term outcomes of the study were published.2 The crude results show that at 8 years the mortality rate was 7.2% lower (p = 0.03) in the early-surgery group. Another interesting data is that rupture risk was four times higher among women than men. We can read this sentence which is a masterpiece of ambiguity: ‘Among patients with a small abdominal aortic aneurysm, we found no long-term difference in mean survival between the early-surgery and surveillance groups, although after 8 years, total mortality was lower in the early-surgery group’.2 Should we imagine that the magic of evidence-based medicine has created that survival, and the mortality is an unrelated phenomenon? The investigators propose three hypotheses to explain this difference and finally reject them to retain a fourth one: lifestyle related to early surgery could underlie the long-term survival benefit from early surgery. This is credible. But, even if the reasons for this higher survival are probably multifactorial, the investigators conceal another simple hypothesis. The UK small AAA trial is a trial of superiority; as previously stated the power of the study does not allow a difference at five years to show. At 8 years the power of the study is higher and a statistically significant superiority becomes ‘evident’ in the early-surgery group. Thus, the conclusion could be that early surgery is the best treatment, but the benefit compared with surveillance is slight. This is the same kind of situation we know in the ‘carotid’ trials for asymptomatic stenosis and symptomatic stenosis between 50 and 69%; a theoretical but thin benefit from surgery.
Internal Validity of the Design

(a) One problem is the credibility of a sharp barrier defined by a duplex measure. Can we really imagine that when an aneurysm grows from 55 to 56 mm the patient’s fate is immediately changed and surgery must be prompted? Another problem is the accuracy of the diameter measured by ultrasounds. In the UK trial, the examiners were selected according to both the observer’s performance and the machine; despite that, the accuracy of the result obtained was plus or minus 2 mm. The question is how to apply the trial conclusion to aneurysms measuring between 53 and 57 mm?

(b) The primary end point of the trial is death, whether or not due to aneurysm rupture. There were 150 deaths in the surveillance group: 17 due to ruptures, one from unknown cause, and the rest attributed to other causes. The autopsy rate in the trial is 29%. Thus 94 primary events are labelled according to the death certificate. Everyone knows that AAA rupture may masquerade a cardiac event.

(c) Among the 527 patients under randomized surveillance, 38 were operated on without fulfilling the criteria for surgery. In the 563 patients undergoing randomized early surgery, 43 patients were not operated on. Globally, these violations of protocol due to cross-over represent 7% of the enrolled patients. In a study where the difference is small or nil, 7% is largely enough to change the conclusions. Although this is a study based on the intention to treat, it would have been interesting to compare the results with a perprotocol study taking into account only the patients complying to the protocol. Looking at the data published it is impossible to establish the consequence of the cross-over patients on the study power.

(d) Another shortcoming of the trial is that conclusions cannot be applied to women. The UK trial comprised only 17% women and 1% in the US trial.

Interpretation and Comments

One has to keep in mind that the cause of death due to rupture may lead to misclassification, so overall mortality could be a better end point than AAA specific mortality in areas with a low autopsy rate. That is what the investigators did, but this basic weak spot should have been discussed and should have led the investigators to temper their conclusion.

The UK small AAA trial as well as the ADAM trial used screening to enrol a substantial number of their participants. The consequence must be a smaller mean size of AAA compared to non-screening areas. If we believe a 5.4 cm AAA has a higher risk of rupture than a 4.0 cm AAA, the overall conclusion may not be true in a non-screened population.

In reply to a critique on subgroup analysis, the investigators strike with arrogance ‘there is no scientific validity for conducting subgroup analysis of the trial data’. Nevertheless, they observe [Ref. 1, Table 2] that there is a lower death risk in the early-surgery group for patients aged under 70 and for AAA between 50 and 55 mm (Table 1). These data suggest a benefit in the early-surgery group for aneurysms above 50 mm and for the younger patients. Question: would an analysis corresponding to this subgroup of patients show the same results?

An important observation is the low rate of ruptures (2.3%) observed during the trial, which may explain the faint role of prophylactic surgery. This low rate is realistic unless there are unknown ruptures in the large group of deaths attributed to other causes. It is interesting to be reminded that during the same time the randomized trial was running, the trial coordinators followed another cohort of 1167 patients ineligible for randomization. The populations of the two groups (1090 randomized patients and 1167 non-randomized patients) were comparable, but the rate of rupture in the non-randomized population was 6.7% (78/1167). The difference may be explained by the scrutiny of surveillance in the UK trial. Question: is this kind of surveillance realistic on a nation-wide scale?

In addition to the burden on the patient, Valentine showed in a study conducted in the Veterans Administration that one third of the candidates for watchful waiting programmes were unable to participate.

Beyond the meanders of statistics we would like to look at things from the patient’s view. The final results of the trial show that after three years 78% of all patients enrolled were operated on, and 84% after six years. During the entire trial, 25 ruptures occurred either for patients under surveillance or for those not operated on under randomized surgery. The non-operated patients were obliged to follow a duplex surveillance, for some of them every three months, with a 70–80% chance to be operated on within a 3–4 year period. We can also say that one patient out of five

<table>
<thead>
<tr>
<th>Table 1. Decrease of death risk by subgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>60–66</td>
</tr>
<tr>
<td>67–71</td>
</tr>
<tr>
<td>Aneurysm diameter (mm)</td>
</tr>
</tbody>
</table>

Eur J Vasc Endovasc Surg Vol 27, April 2004
will have the chance to live for 4 years with a duplex every 3 months, in fear of rupture, and being almost certain to be operated at the end of the story.

And finally, here is the investigator’s point of view. The first objective of the UK small AAA trial was ‘to determine whether early elective surgery or a period of US surveillance provides the better management of small asymptomatic AAA…’. One wonders whether the investigator’s attitude is neutral or if their inmost thoughts are to demonstrate that early surgery must be condemned. They righteously defend the surprisingly high mortality rate due to a pragmatic attitude in their approach to recruitment and fitness for surgery. If the ‘no selection’ of surgeons is pragmatic, can we say that the selection of duplex specialists and the frequency of surveillance are pragmatic? Do all the duplex specialists in the country obtain the same accuracy of plus or minus 2 mm? Are all the patients screened with a 40–55 diameter aneurysm as compliant as the ones enrolled in the trial after having signed an informed consent?

According to objective number one of the trial, the investigators assess the quality of life and conclude there was a benefit after 12 months in the early surgery group. Immediately after that, they add this comment ‘for families who have lost a loved one after surgery, quality of life may have deteriorated substantially’. Did they think of the families who have lost a loved one after rupture?

In a correspondence, the investigators say ‘this may be a British prerogative, better to understate than overstate’. In this case, to claim clear-cut conclusions from this trial is clearly an overstatement. As the ADAM investigators say ‘arbitrarily setting a single threshold diameter for elective AAA repair in all patients is naïve’. We must thank and congratulate the UK small aneurysm trial investigators for their splendid job and for the large amount of knowledge with which they provided us. But we would advise them to be more modest and change their diktats for more pragmatic conclusions. Diameter measurement, especially with duplex, is always an approximation; the crude number must be replaced in the context of the patient. Surveillance is an acceptable alternative for AAA between 40 and 55 mm, probably the best alternative for high risk patients. Surgery is the more reasonable solution for good risk patients with an AAA > 50 mm.

Acknowledgements
We thank Nicolas Valério (MD; Department of Vascular Surgery) and Roch Giorgi (MD; Department of Biostatistics) who helped us to analyse the data and prepare the manuscript.

References

Accepted 21 January 2004