Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding?

Dean Fergusson\textsuperscript{a,b}, Kathleen Cranley Glass\textsuperscript{b,c}, Brian Hutton\textsuperscript{a} and Stan Shapiro\textsuperscript{b,c,d}

\textbf{Background} Aprotinin is a serine protease inhibitor used to limit perioperative bleeding and reduce the need for donated blood transfusions during cardiac surgery. Randomized controlled trials of aprotinin evaluating its effect on the outcome of perioperative transfusion have been published since 1987, and systematic reviews were conducted in 1992 and 1997.

\textbf{Methods} A systematic search was conducted for all RCTs of aprotinin that used placebo controls or were open-label with no active control treatment. Data collected included the primary outcome, objective of each study, whether a systematic review was cited or conducted as part of the background and/or rationale for the study and the number of previously published RCTs cited. Cumulative meta-analyses were performed.

\textbf{Results} Sixty-four randomized, controlled trials of aprotinin were found, conducted between 1987 and 2002, reporting an endpoint of perioperative transfusion. Median trial size was 64 subjects, with a range of 20 to 1784. A cumulative meta-analysis indicated that aprotinin greatly decreased the need for perioperative transfusion, stabilizing at an odds ratio of 0.25 ($p < 10^{-6}$) by the 12th study, published in June of 1992. The upper limit of the confidence interval never exceeded 0.65 and results were similar in all subgroups. Citation of previous RCTs was extremely low, with a median of 20\% of prior trials cited. Only 7 of 44 (15\%) of subsequent reports referenced the largest trial ($N = 1784$), which was 28 times larger than the median trial size.

\textbf{Conclusions} This study demonstrates that investigators evaluating aprotinin were not adequately citing previous research, resulting in a large number of RCTs being conducted to address efficacy questions that prior trials had already definitively answered. Institutional review boards and journals could reduce the number of redundant trials by requiring investigators to conduct adequate searches for prior evidence and conducting systematic reviews. \textit{Clinical Trials} 2005; 2: 218–232. www.SCTjournal.com

\section*{Introduction}

The ostensible purpose of randomized controlled trials (RCTs) is to answer an unsettled question about an intervention’s efficacy. This paper uses the example of RCTs of aprotinin treatment to address the issue of how much experimentation is enough, and if better procedures are needed to prevent RCTs...
from being initiated or conducted when the question they are addressing has already been answered.

Aprotinin is a serine protease inhibitor used to limit perioperative bleeding and thus reduce the need for allogeneic (donated) red blood cell transfusions [1]. A meta-analysis of 16 trials of aprotinin in cardiac surgery published in 1994 [2] concluded that it was highly effective in reducing the proportion of patients requiring a transfusion (odds ratio 0.23; 95% confidence interval, 0.16 to 0.33). Another meta-analysis was published in 1997 by one of the authors (DF) [3]. This review of 45 randomized clinical trials further affirmed the effectiveness of aprotinin (odds ratio 0.31; 95% confidence interval, 0.25 to 0.39). While conducting the meta-analysis it became obvious that randomized controlled trials (RCTs) had continued to be proposed, funded, conducted, and published well after the effectiveness of aprotinin had been established.

An RCT is permissible when there is no scientific consensus about the relative efficacy of two competing interventions. Freedman referred to this state of uncertainty in the expert community as clinical equipoise [4]. Clinical equipoise provides a justification for randomising individuals to competing therapies or to placebo. Clinical equipoise must be based on awareness of the existing medical evidence, which in turn requires the organized study of that evidence. This should be presented to colleagues, Institutional Review Boards (IRBs), funding and regulatory agencies, journal editors and peer reviewers, and prospective participants as part of the justification for a given RCT. In light of the obligation to justify a claim of clinical equipoise between treatment options before proceeding with a trial, one of the first questions to be asked is whether trialists systematically reviewed the prior literature.

Methods

An unrestricted Medline and EMBASE literature search was conducted for the dates January 1966 to March 1997 with the text word aprotinin to identify randomized controlled trials in cardiac surgery. An updated systematic literature search with an RCT filter [5] was conducted to identify all cardiac surgery RCTs indexed in Medline after 1996. Only randomized trials that described the proportion of patients receiving at least one unit of allogeneic red blood cells were eligible. Studies were included regardless of whether they were full publications, abstracts or letters to the editor; or were published in a language other than English. All RCTs had to be either placebo controlled or open-label with no active control. Duplicate publications, publications without data on the proportion of patients transfused, and single-centre publications that were part of a multi-centre publication were excluded. Pseudo-randomized trials (e.g., randomized by birthdate), noncontrolled trials, review articles, and observational studies were excluded. RCTs with an active comparator and no open-label or placebo control were excluded.

Two reviewers, independently, assessed each citation for eligibility and a total of 62 publications representing 64 trials met full inclusion criteria [6–67]. In addition to the 45 trials from a previous meta-analysis [3,6–45,47,48,50,51], the updated literature search identified 19 further trials published between 1996 and 2004 [46,49,52–67].

Data collected from each trial included the objective, patient characteristics, whether a systematic review was conducted as part of the background and/or rationale for the study, number of subjects, publication date, dates of study enrolment, and all previously published randomized trials and systematic reviews cited. In addition, trial quality was assessed using a published, validated quality scale [68]. If the study stated more than one objective without stating which one was the primary objective, all were considered as primary. A cumulative meta-analysis that produces an updated measure of effect by statistically pooling studies after each new study is completed was performed with all 64 trials. Due to the lack of reporting of dates of randomization, we used the publication date as the study completion date.

To elucidate possible reasons for the continued use of placebo or open-label control arms, subgroup cumulative analyses were performed stratified by methodological quality and patient characteristics. To evaluate the effect of trial quality upon the results of these meta-analyses, the quality scale was used [68] along with a subgroup analysis of open-label versus placebo controls. The trial quality scale assesses quality based on randomization, blinding, and the description of withdrawals. The highest possible score is five; the lowest is zero. A score equal to or greater than three was considered good and less than three was considered poor quality. This judgment is consistent with the original publication [68]. Each trial was evaluated independently by two individuals, with differences resolved by either consensus or independent evaluation of a third party. Outcome data from each trial (proportion of subjects requiring at least one unit of allogeneic red blood cell transfusion) was analysed using software (Meta-Analyst977, J Lau and T Chalmers) with a random-effects model. Effect sizes are presented as odds ratios (OR) with 95% confidence intervals. An OR of 1 suggests no difference between intervention and control; an OR <1 suggests that fewer subjects
Results

Figure 1 presents the date of publication, sample size and odds ratio for each of the 64 trials. Overall, there were 8040 subjects entered in the 64 trials. Median trial size was 64 subjects (range 20–1784) with 46 of 64 (72%) enrolling fewer than 100 patients. The largest trial consisted of 1784 patients and was published in 1992. Figure 2 illustrates the published randomized trials cited in each of the 64 trials. Dates of study enrolment were reported in 15 of the 64 trials.

The cumulative meta-analysis presented in Figure 3 indicates that a clinically significant result was achieved in the very first trial of 22 patients (OR 0.03, 95% CI: 0.00–0.56). After the 12th study, the cumulative effect estimate stabilizes in the range of 0.25–0.35. Variability around this estimate narrows as the cumulative sample size increases. Throughout the cumulative meta-analysis, the upper limit of the confidence interval never crosses 0.65.

Results of the subgroup analysis for good quality and placebo-controlled trials are presented in Figure 4a and b. By including only the studies assessed as good quality, a highly clinically and statistically significant association is evident by 1990 after the third such trial (OR 0.09, 95% CI 0.02–0.54). A further 31 good quality randomized controlled clinical trials were published between 1990 and 2001. Results are similar for sub group analyses of placebo-controlled studies.

The objectives of each of the trials were examined to identify possible justifications for conducting further trials. All 64 publications stated an objective with blood loss or transfusion requirements mentioned as a primary objective or outcome in 49 (77%) of the trials. Of the 64 trials, 53 stated whether the patient population included primary (38 trials), repeat (five trials), or a combination of primary and repeat surgery (10 trials) cases. Separate cumulative meta-analyses indicate that effectiveness in each of the three surgical groups was established in the early 1990s (Figure 5). Of the 64 trials, 51 provided information on aspirin use (Figure 5). For the 24 trials that enrolled a proportion of patients who were taking aspirin at the time of surgery, the upper limit of the confidence interval did not cross 0.62 after the first trial. For trials enrolling patients exclusively taking aspirin at the time of surgery, the cumulative effect size became nonsignificant after the fourth of five trials. After the fifth trial, the overall effect was an odds ratio of 0.39 (95%CI: 0.17–0.89).

Discussion

To be ethical, clinical research must be valuable [69]. To be of value, a trial must add to current knowledge. An integral step in evaluating the evidence is to conduct a systematic review of the literature. A systematic review or meta-analysis refers to an overview of the literature conducted in a well-defined, systematic and thorough manner, be it quantitative or qualitative. Beyond justifying a research question, a systematic review serves three purposes: 1) trialists become aware of the full extent of the literature related to their research question; 2) it provides a transparent, traceable path of due diligence for research review boards, journal editors, readership and, ultimately, prospective patients; and 3) it provides trialists a comprehensive list of clinical, design, and statistical issues that may be relevant for the design or interpretation of their proposed trial.

A multi-country survey conducted between 1995 and 1997 found that drugs to minimize peri-operative bleeding and transfusion requirements were used in a large proportion of hospitals [70]. Of these drugs, aprotinin was used in the greatest proportion of hospitals. While it is arguable whether or not aprotinin was or should be a standard of care at cardiac surgery centres due to issues of costs, safety and the availability of other agents, the effectiveness of aprotinin at reducing transfusion requirements and blood loss, the ostensible focus of the vast majority of RCTs of aprotinin, has been well established since the mid-1990s [2,3]. More importantly, the use of antifibrinolytic pharmacological agents in cardiac surgery, especially when substantial blood loss or transfusion requirement is expected, was also well established.
Randomized controlled trials of aprotinin in cardiac surgery

Figure 1 Proportion of patients transfused.
Our study illustrates that trialists evaluating aprotinin were not adequately citing previous trials nor conducting systematic reviews to support the need for additional trials. It seems reasonable to conclude that there remains a barrier between published evidence-based medicine (clinical trials and systematic reviews) and the review of this evidence in considering whether further trials are warranted. None of the 64 aprotinin trials reported that a systematic review or meta-analysis had been conducted and only two (3%) trials referenced published systematic reviews. Instead, selective early trials, especially the initial trial, were cited to support the objective of the study. Thirty-six of 63 subsequent trials (57%) referenced the first published systematic reviews. Instead, selective early trials, especially the initial trial, were cited to support the objective of the study. Thirty-six of 63 subsequent trials (57%) referenced the first published systematic reviews. Instead, selective early trials, especially the initial trial, were cited to support the objective of the study.

Figure 2: Citations of previous studies.

Figure 6 demonstrates a profound and troubling gap in available versus cited publications. The largest trial (by an order of magnitude) was not cited by 37 of 44 trials published more than a year later. All 62 publications representing 64 trials were easily identifiable through Internet literature search portals and all trials, save one, were indexed on Medline within weeks of publication. A simple literature search using the National Library of Medicine's universally accessible and free of charge PubMed world wide web portal with the term “aprotinin” restricted to the publication type “randomized controlled trials” identified 58 of the 62 publications (94%) and a PubMed search with the terms “aprotinin” and “random” identified 51 (82%) publications. Moreover, the vast majority of articles were published in top-tier cardiothoracic specialty journals.

Clarke and Gotzsche have addressed the important issue of citing previous research [71,72]. They concluded that researchers do not satisfactorily address nor do journals adequately reflect the...
### Odds Ratios with 95% Confidence Intervals

<table>
<thead>
<tr>
<th>Ref #</th>
<th>Year of Publication</th>
<th># Pts</th>
<th>Favours Aprotinin</th>
<th>Favours Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Dec-87</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Mar-89</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Apr-89</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Sep-90</td>
<td>219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Oct-90</td>
<td>257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Dec-90</td>
<td>296</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Jun-91</td>
<td>376</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Sep-91</td>
<td>396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Dec-91</td>
<td>455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Apr-92</td>
<td>486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Jun-92</td>
<td>601</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Jun-92</td>
<td>2385</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Jun-92</td>
<td>2445</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Nov-92</td>
<td>2495</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Dec-92</td>
<td>2664</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Jan-93</td>
<td>2754</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Jul-93</td>
<td>2795</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Aug-93</td>
<td>3005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Dec-93</td>
<td>3044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Jan-94</td>
<td>3146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26a</td>
<td>Feb-94</td>
<td>3201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26b</td>
<td>Feb-94</td>
<td>3342</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Feb-94</td>
<td>3396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Apr-94</td>
<td>3475</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Jul-94</td>
<td>3575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Aug-94</td>
<td>3668</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Aug-94</td>
<td>3724</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Oct-94</td>
<td>3822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Oct-94</td>
<td>3854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Dec-94</td>
<td>3882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Dec-94</td>
<td>4047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Feb-95</td>
<td>4147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Feb-95</td>
<td>4210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Feb-95</td>
<td>4240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Apr-95</td>
<td>4338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Jun-95</td>
<td>4382</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Jun-95</td>
<td>4420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Sep-95</td>
<td>4450</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Oct-95</td>
<td>4548</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Oct-95</td>
<td>4578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Oct-95</td>
<td>4832</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>May-96</td>
<td>4882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Jul-96</td>
<td>4975</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Aug-96</td>
<td>5023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Aug-96</td>
<td>5135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Oct-96</td>
<td>5326</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>Dec-96</td>
<td>5970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>Jan-97</td>
<td>6008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>Jan-97</td>
<td>6060</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Aug-97</td>
<td>6227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>Sep-97</td>
<td>6333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>Dec-97</td>
<td>6376</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57a</td>
<td>Oct-98</td>
<td>6442</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57b</td>
<td>Oct-98</td>
<td>6507</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Nov-98</td>
<td>7303</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>Aug-99</td>
<td>7360</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>Sep-99</td>
<td>7510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>Mar-00</td>
<td>7593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>Dec-00</td>
<td>7677</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>Dec-00</td>
<td>7697</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>Jan-01</td>
<td>7897</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>Sep-01</td>
<td>7952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>Sep-01</td>
<td>8011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>Jun-02</td>
<td>8040</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0.11 (0.03, 0.38)  
0.22 (0.09, 0.52)  
0.28 (0.20, 0.38)  
0.29 (0.23, 0.38)  
0.30 (0.24, 0.38)  
0.33 (0.26, 0.41)  
0.34 (0.29, 0.41)

**Figure 3** Cumulative meta-analysis of all RCTs.
Gotzsche notes that identifying previous studies by simply scanning bibliographies of a convenient sample of publications can produce a biased sample of articles [72]. This is referred to as reference bias. The patterns in Figure 2 provide a strong suggestion of this: trials referenced in one paper were often picked up in subsequent papers, but those that were missed early remained largely uncited. This is evidence that researchers were not conducting their own independent reviews of the literature, but were instead depending on previous incomplete searches by others.

Unfortunately, too few trial publications (23%) provided dates of patient accrual that would have given a more accurate depiction of publications available to investigators at time of study commencement. Of those reporting accrual dates, the vast majority accrued over a one-year period with the trial publication one year after patient enrolment ended. Even accounting for substantial lag time, Figures 2 and 6 clearly demonstrates that available evidence was not being evaluated in a systematic manner prior to the start of these trials. This makes it inevitable that trials were conducted that did not need to be initiated.

Our present study focused on identifying all randomized trials that reported the proportion of patients transfused, as this outcome reflects the 1993 Food and Drug Administration’s approval indication. With respect to their objectives and outcome measures, the trials were quite homogenous. Of the 64 trials, 15 provided a primary outcome or objective other than blood transfusion or blood loss. Nine trials evaluated graft patency or myocardial infarction as a primary objective, but no trial had the primary objective of assessing allergic reactions, mortality, or other serious thrombotic events. As for homogeneity with respect to patient populations, separate cumulative analyses were carried out for primary surgery patients only, repeat surgery patients only, combination of primary and repeat surgery patients, patients all on aspirin, some patients on aspirin, no patients on aspirin, and a

---

**Figure 4** Cumulative meta-analysis of placebo-controlled trials and cumulative meta-analysis of good quality trials.
range of aprotinin doses from ultra low to high. The effectiveness of aprotinin was established early and remained consistent in all the above categories of patients except for trials conducted exclusively in patients taking aspirin at the time of surgery. Given the estimates provided in Figure 1, it is not surprising that the effectiveness of aprotinin remains consistent across different clinical subgroups.

The 64 RCTs examined here were not the only studies conducted and published on this topic. Numerous randomized trials not examining the outcomes studied here were conducted, as were numerous pseudo- or nonrandomized trials, active-controlled and uncontrolled trials of aprotinin. Indeed, between 1997 and 2004, 61 randomized controlled trials of aprotinin in cardiac surgery were identified in our systematic literature search, only 19 of which met eligibility for the present study.

Despite the substantial efficacy evidence that has accrued on this treatment, trials of aprotinin in cardiac surgery continue. One could argue that not all relevant outcomes have been sufficiently studied, such as serious adverse events (e.g., thrombotic events, mortality, myocardial infarction). However, to examine these endpoints trials of much greater sample size would have to be conducted and trialists would have had to provide evidence that the risk of harm from administering aprotinin outweighed the risks of not receiving aprotinin in the placebo group. Other unsettled questions could be that certain patient subsets may respond differently to different administration schedules and/or doses of aprotinin. But those questions do not require the use of a placebo or open-label arm. If dosing is the primary objective the control arm should receive the already established effective dose(s) as outlined in the product monograph.

Another possible motivation for conducting a trial is to gain local experience with an intervention before it is adopted. Local factors such as surgical expertise, transfusion thresholds, and various co-interventions can alter the effectiveness of
aprotinin in a particular setting. However, employing a randomized controlled trial for the purpose of gaining local experience is less than ideal from an ethical or pragmatic standpoint.

IRBs are established to ensure that research involving humans meets the established ethical requirements, which are summarized by Emanuel: value, scientific validity, fair subject selection, favourable risk-benefit ratio, independent review, informed consent and respect for potential and enrolled subjects [73]. To this end, IRBs must critically evaluate the background and rationales provided by the investigators. Some have questioned whether IRBs have duly fulfilled their role, citing recent trials of established effective treatments using placebo control. Others have suggested that the performance and accountability of IRBs would be improved by requiring trialists to submit systematic reviews in support of their application [74]. This requirement affirms the responsibility of both IRB and investigator to adequately consider the extant evidence base when assessing the need for a study, and it allows the IRB to consider the evidence in an unbiased and transparent manner. The results of our study show the potential consequences of not having such a requirement.

As the single largest disseminator of research results, journals have a responsibility for ensuring they publish only scientifically and ethically valid and valuable research. Assessing whether clinical equipoise was present at the start of the trial must be part of the editorial calculus. One suggestion would be that the CONSORT statement [75] be amended to require authors to explicitly state whether a systematic review was conducted to support a state of clinical equipoise. The systematic review can either be an original undertaking or an update of a previous systematic review.

Over the past century and a half we have made considerable progress in transforming the art of medicine into the science of medicine [76]. However, systematizing the review of available evidence remains an area where we are moving more slowly than desirable. How do we prevent unnecessary trials? Requiring investigators to conduct more organized reviews of the evidence using
well-established standards and scientific methods of systematic reviewing, with the diligence of this effort assessed by IRBs and journals, would be a constructive step in the right direction.

Acknowledgements

The authors thank Dr Paul Hébert and Dr Leon Glass for the helpful comments and suggestions in the preparation of this manuscript.

References


62. Santamaria A, Mateo J, Oliver A et al. The effect of two different doses of aprotinin on hemostasis in...


Discussion

Comment

The scandalous failure of science to cumulate evidence scientifically

Iain Chalmers*

The article by Dean Fergusson and his colleagues [1] in this issue of the journal [p] is the most recent evidence of an ongoing scandal in which research funders, academia, researchers, research ethics committees and scientific journals are all complicit. New research should not be designed or implemented without first assessing systematically what is known from existing research [2,3]. The failure to conduct that assessment represents a lack of scientific self-discipline that results in an inexcusable waste of public resources. In applied fields like health care, failure to prepare scientifically defensible reviews of relevant animal and human data results not only in wasted resources but also in unnecessary suffering and premature death [4–11].

Fergusson and his colleagues [1] have used the technique of cumulative meta-analyses of randomized trials, pioneered by Tom Chalmers and his colleagues more than a decade ago [4,5], to analyse 64 trials assessing the effect of aprotinin on perioperative blood loss, as judged by the use of blood transfusion. They show, as they did eight years ago [12], that placebo controlled trials of aprotinin have continued to be done long after strong evidence has accumulated showing that the drug substantially reduces the use of blood transfusion.

In addition to this litany of unnecessary, and therefore unethical research, Fergusson and his colleagues present an analysis of the extent to which authors of the reports of the 64 trials cited relevant earlier trials. Their shocking findings are summarized in Figures 2 and 6: between 1987 and 2002 the proportion of relevant previous reports cited in successive reports fell from a high of 33% to only 10% among the most recent reports. Furthermore, only seven of 44 subsequent reports referenced the report of largest trial (which was 28 times larger than the median trial size); and only two of the reports referenced systematic reviews of these trials published in 1994 and 1997.

---

*aIain Chalmers is editor of The James Lind Library (www.jameslindlibrary.org), a web-based resource containing material about the evolution of fair tests of medical treatments, and co-ordinator of the secretariat for the James Lind Alliance (www.lindalliance.org), a coalition of patients and clinicians collaborating to confront important uncertainties about the effects of treatments. He was director of the UK Cochrane Centre between 1992 and 2002, and director of the National Perinatal Epidemiology Unit between 1978 and 1992.*
This is simply the latest example of the consequences of a lack of scientific and ethical self-policing among researchers and those who fund their activities. But what of the research ethics committees that approved these studies? Ten years have passed since ethics committees were challenged publicly to recognize that they were behaving unethically by not taking steps to assure that they were approving only necessary research [6], yet there is very little evidence that they have taken this challenge seriously [13,14].

And what were the editors of journals doing accepting reports of redundant research for publication? Had any of them taken seriously the proposal that systematic reviews should be used by editors and peer reviewers to judge submitted manuscripts in the context of related, previous studies [15]?

I propose that the research ethics committees and journals who approved and published studies of aprotinin after 1990 should be invited to send Clinical Trials their comments on the paper by Fergusson and his colleagues. Not only would this help to show ethics committees and editors how they are failing patients and the public in this domain, but publication of these comments should help to uncover some of the academic, commercial and practical pressures that are leading to this indefensible situation.

As the paper emphasizes, science is meant to be cumulative, but many scientists are not cumulating scientifically – and those who can call researchers and academia to account are failing to do so. Not only are most new studies not designed in the light of systematic reviews of existing evidence, new evidence is only very rarely reported in the context of systematic reviews of existing evidence, and reported setting the new evidence “in the light of the totality of the available evidence” [22], thus making clearer to readers what contribution – if any – new studies have made to knowledge.

The idea that new research results should be set in context has existed for well over a century [3]. In 1884, in his Presidential Address to the meeting of the British Association for the Advancement of Science in Montreal, Lord Rayleigh, Professor of Physics at the University of Cambridge, noted that “the work which deserves, but I am afraid does not always receive, the most credit is that in which discovery and explanation go hand in hand, in which not only are new facts presented, but their relation to old ones is pointed out” [19].

The scientific and ethical consequences of academia’s failure to take research synthesis sufficiently seriously in biomedical and clinical research [20] are that patients (and the public more generally) suffer directly and indirectly; policymakers, practitioners, and patients have inadequate information to guide their choices among alternatives; and limited resources for health care and new research are used inefficiently. Those who wield power within academia should either publicly defend their failure to take effective action [21], or act more forcefully to change this unacceptable state of affairs.

This paper by Fergusson and his colleagues compellingly demonstrates why all new research – whether basic or applied – should be designed in the light of scientifically defensible syntheses of existing research evidence, and reported setting the new research “in the light of the totality of the available evidence” [22], thus making clearer to readers what contribution – if any – new studies have made to knowledge.

References

12. Laupacis A, Fergusson D. Drugs to minimize perioperative blood loss in cardiac surgery:


Comment

John G. Augoustides MD, FASEa and Lee A. Fleisher MD, FACCb

How many randomized controlled clinical trials (RCTs) are required to document that aprotinin improves hemostasis after cardiac surgery? Clinical equipoise, real doubt about drug efficacy, is violated when RCTs are conducted even after drug efficacy has been proven. When clinical equipoise is violated in an RCT, patients may be unethically randomized to placebo and denied proven benefit from the “study drug”.

Fergusson et al. present a comprehensive meta-analysis (64 RCTs: n = 8040) that details the extent of redundant aprotinin RCTs conducted after the hemostatic efficacy of aprotinin in cardiac surgery was established. An RCT to evaluate drug efficacy is redundant if efficacy has already been proven. In 1997 Fergusson co-authored an extensive meta-analysis (45 RCTs: n = 5808) that quantified the hemostatic efficacy of aprotinin in cardiac surgery showing a significantly decreased allogeneic blood exposure (odds ratio 0.31, 95% confidence interval 0.25–0.39; P < 0.0001) and significantly decreased reoperation rate for bleeding (odds ratio 0.44, 95% confidence interval 0.27–0.73; P = 0.001) [1].

In this follow-up meta-analysis, focused more on aprotinin RCT design rather than efficacy, Fergusson et al. demonstrate and quantify the degree of RCT redundancy. Substantial aprotinin efficacy was observed in the very first of 64 RCTs; even when the analysis was restricted to those trials deemed of highest quality, 31 aprotinin RCTs were published after 1990, when aprotinin efficacy should have been concluded from the first three trials. Redundancy persisted in the subgroup analyses stratified by aspirin exposure and primary and repeat cardiac surgery. As stressed by these investigators and by many who have preceded them [2–5], a systematic literature review preceding a trial should prevent RCT redundancy.

What are the possible explanations for this apparent tremendous redundancy in aprotinin RCTs? The first possible may be that the most cost-effective hemostatic aprotinin dosage regimen remained undefined, an issue because high-dose aprotinin is expensive. Perhaps ongoing RCTs were required to delineate the lowest dose of aprotinin that produced adequate hemostasis [6–9].

A second possible explanation for the seeming excess RCTs is that they were mounted to better quantify the balance between the benefit on the bleeding outcome and the risk of vascular graft thrombosis, which was a major controversy [10,11]. Although the authors claim this would require larger RCTs than were conducted, this may merit further inquiry.

A third possible explanation is that aprotinin’s biologic effects beyond hemostasis justified evidence-based evaluation, such as in vitro anticoagulant properties [12], platelet-sparing effects [13], anti-inflammatory properties [14,15], and organ protection properties [16]. It is doubtful whether such endpoints would justify RCTs in the face of proven clinical efficacy, but it would help us better understand the phenomenon observed here to know how many of the protocols cited such issues as part of the trial justification.

aJohn G. Augoustides is an Assistant Professor of Anesthesia, University of Pennsylvania School of Medicine. He is a Fellow of the American Society of Echocardiography. He practices as a cardiovascular anesthesiologist, and has a special interest in thoracic aortic procedures, including aprotinin.
bLee A. Fleisher is the Robert D. Dripps Professor and Chair of Anesthesia as well as Professor of Medicine at the University of Pennsylvania School of Medicine. He is an associate scholar of the Center for Clinical Epidemiology and Biostatistics at the University.
A fourth possible explanation is that redundant aprotinin RCTs were conducted to improve market share and diffuse information about the effectiveness of aprotinin into regions that did not use it. These motives for RCTs are less than ideal, and represent suboptimal clinical practice, but the ethical argument becomes more complex if the routine care in the study center did not include the drug.

In summary, this high-quality meta-analysis demonstrates significant redundancy in aprotinin RCTs within the cardiac surgical population, including subgroups exposed to aspirin, primary and repeat surgery. A full understanding of the reasons for this redundancy may require further investigation with respect to the four justifications outlined here. These investigations might reduce what we judge to be the degree of aprotinin RCT redundancy, but will probably not eliminate it.

In the future we can anticipate more aprotinin RCTs, addressing clinical concerns such as 1) safety in specific cardiac surgical subsets (off-pump coronary bypass, deep hypothermic circulatory arrest); 2) platelet protection (patients exposed to aspirin and clopidogrel); and 3) organ protection (the brain and heart, in particular). Each RCT is justified as long as clinical equipoise is respected and the hypothesis is clearly formulated a priori, based on a thorough literature review. This meta-analysis highlights how essential these steps are for responsible RCT design. We will need to follow them to ethically advance our understanding, in an evidence-based fashion, of ways to improve perioperative outcomes for our patients.

References