The Need for Large Clinical Studies in Perioperative Medicine

In recent decades, substantial advances in surgery have improved disease treatment and patients’ quality of life. Consequently, increasing numbers of patients are having surgery. A recent study that used surgical data from 56 countries suggests that more than 230 million major surgical procedures are undertaken annually around the world.\(^1\)

Despite its benefits, surgery is associated with major complications. For example, worldwide approximately 5 million adults annually suffer a major perioperative vascular complication (i.e., vascular death, nonfatal myocardial infarction, nonfatal cardiac arrest, and nonfatal stroke) in the first 30 days after noncardiac surgery.\(^2\) This incidence, remarkably, is similar to the annual global incidence of new cases of human immunodeficiency virus.\(^3\) Despite the magnitude of the problem, there is not a single established effective and safe intervention to prevent major vascular complications after noncardiac surgery.\(^4\) Furthermore, our ability to predict these events is limited. Other complications (e.g., respiratory) are also common and cause substantial morbidity and mortality.\(^5\)

**Perioperative Research**

Unlike cardiology, large clinical studies remain uncommon in perioperative medicine.\(^6,7\) Further, there has been a tendency to believe the results of small perioperative clinical studies, especially when they demonstrate statistically significant results. This position is supported by the fact that perioperative guideline committees recommended β-blockers to patients undergoing noncardiac surgery for a decade based upon the results of small trials demonstrating implausibly large treatment effects.\(^5,9\) Unfortunately, the results from small clinical studies are not reliable and are often wrong, as history has shown in other areas of medicine.\(^10\)

**Historical Lessons**

Cellular, physiologic, genetic, and biochemical studies are clearly required to understand human biology and pathologic process. They are, however, not designed to guide patient care, preventive strategies, or health policy decisions. For example, as outlined in figure 1: 1) some drugs that improve cardiac inotropy or provide afterload reduction and increased exercise capacity\(^11\) – and would thus be expected to decrease mortality – actually increase mortality\(^12,13\); 2) large randomized controlled trials (RCTs) do not support the benefits of antioxidants that were strongly suggested in biomedical research\(^14,15\); 3) all drugs that improve atherogenic lipid profiles and atheromatous plaque burden\(^16\) do not improve patient-important cardiovascular outcomes\(^17,18\); and 4) some drugs that decrease atheromatous plaque\(^19\) do not reduce the risk of major cardiovascular events.\(^20\)

A rich body of literature has clearly shown that questions of diagnosis, prognosis, and therapy are most reliably answered by large clinical studies including observational studies (i.e., cohort and case-control studies) and RCTs of both individual-based interventions and population-based strategies. Observational studies establish independent risk factors and prediction and prognostic models for a given clinical state/disease and identify potential prevention or treatment interventions for further testing. RCTs are, however, the...
most reliable method to determine the clinical response to an intervention.21

There is a common misunderstanding that “discovery” is only done in small studies or basic sciences. The reality is, however, that large observational studies and trials are both platforms for discovery research and evaluation research, as well as application research. For example, preventing myocardial infarction and stroke with an angiotensin-converting-enzyme inhibitor was not based initially on laboratory or animal science, but on clinical observations in large clinical studies, which in turn led to both basic science and large trials.22–24 As a result of this research, the use of angiotensin-converting-enzyme inhibitor therapy now prevents an estimated 1 to 2 million premature vascular events annually, with concomitant enormous cost savings. Similarly, perioperative β-blockade was not known to increase stroke risk until a large international RCT demonstrated a perioperative β-blocker doubled the risk of stroke.25

Why Is It Necessary to Have Large Populations Participate in RCTs?

There are two fundamental reasons why we need large sample sizes in RCTs to produce reliable results. First, the pathophysiology of most disease states is multifactorial, and given that most interventions target just one or two risk factors, it

Fig. 1. Examples of large trials refuting evidence from biomedical or physiologic studies. (A) Milrinone in a before-after study of 12 patients increased maximum exercise capacity by 22% (P < 0.001).11 Milrinone in a randomized controlled trial (RCT) of 1,088 patients, in which 295 patients died, demonstrated that milrinone increased mortality by 28% relative to placebo (P = 0.038).12 (B) Vitamin E in an experimental study of 36 rabbits demonstrated that vitamin E supplement attenuated atherosclerotic lesions by 32% relative to no vitamin E supplement (P < 0.05).15 Vitamin E in a RCT of 9,541 patients, in which 1,511 patients developed a major cardiovascular (CV) event (myocardial infarction, stroke, or CV death), showed that vitamin E did not prevent major CV events (relative risk 1.05; P = 0.33).14 (C) Hormone therapy in a RCT of 55 monkeys demonstrated that hormone therapy decreased coronary plaque by 55% relative to no hormone therapy (P = 0.05).16 Hormone therapy in a RCT of 16,608 women without CV disease, in which 1,240 developed a CV event (myocardial infarction, stroke, coronary revascularization, or venous thromboembolic event), showed that hormonal therapy increased the risk of a CV event (hazard ratio 1.22; 95% CI, 1.09–1.36).17 (D) Azithromycin in a RCT of 19 rabbits infected with Chlamydia pneumoniae demonstrated that azithromycin resulted in less atheromatous maximal intimal thickness by 64% relative to no azithromycin (P < 0.025).19 Azithromycin in a RCT of 7,747 patients with a history of myocardial infarction and Chlamydia pneumoniae infection, in which 1,130 suffered a major CV event (myocardial infarction, coronary revascularization, hospitalization for angina, or death), showed that azithromycin did not prevent major CV events (relative risk 7%; P = 0.23).20
is unlikely that any particular intervention will have more than a moderate size treatment effect (i.e., relative risk reductions in the range of 25%). For example, although statin therapy is extremely effective at decreasing cholesterol levels, these drugs decrease cardiovascular events by about 25%, not 75%. The reason for this moderate treatment effect is that many patients on statins continue to smoke, have diabetes mellitus, poorly controlled hypertension, and obesity. Large RCTs are required to reliably detect the plausible yet important moderate size effects that can have a substantial impact on health.

Given the plethora of perioperative triggers (e.g., surgical trauma, pain, bleeding, fasting) that initiate potentially harmful states (e.g., inflammatory, hypercoagulable, stress, hypoxic), it is only realistic to expect moderate-sized treatment effects from the vast majority of potential perioperative interventions. Consistent with this point, the initial report of an unrealistically large reduction in myocardial infarction with perioperative β-blockade (relative risk reduction 100% in a trial of 112 patients) was subsequently shown to be substantially exaggerated when a trial of 8,351 patients demonstrated perioperative β-blockade resulted in a relative risk reduction of 27% for myocardial infarction. Large trials are needed to ensure reliable estimates of treatment effects.

The second reason why we need large RCTs relates to a common error in interpreting evidence and the confidence clinicians have in a study based upon the statistical significance of the results, despite the size of the study and number of events. Consider two hypothetical RCTs each evaluating the effects of an investigational drug versus placebo in patients at risk of a perioperative myocardial infarction. Both trials use identical high-quality methodology, including concealment of randomization, blinding, complete patient follow-up, and intention-to-treat analyses. The first trial randomizes 100 patients to receive investigational drug A and 100 patients to receive placebo, and fewer patients assigned the investigational drug suffer a perioperative myocardial infarction (1 vs. 9 patients, \( P = 0.02 \), Fisher exact test, table 1). The second trial randomizes 4,000 patients to receive investigational drug B and 4,000 patients to receive placebo, and fewer patients assigned the investigational drug suffer a perioperative myocardial infarction (200 vs. 250 patients, \( P = 0.02 \), table 1).

Table 1. Results from a Hypothetical Set of Trials

<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Treatment A</th>
<th>Placebo</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>(n = 100)</td>
<td>(n = 100)</td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>1</td>
<td>9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial 2</th>
<th>Treatment B</th>
<th>Placebo</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>(n = 4,000)</td>
<td>(n = 4,000)</td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>200</td>
<td>250</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 2. Results Demonstrating Fragility of Results in a Small Trial

<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Treatment A</th>
<th>Placebo</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>(n = 100)</td>
<td>(n = 100)</td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>3</td>
<td>9</td>
<td>0.13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial 2</th>
<th>Treatment B</th>
<th>Placebo</th>
<th>( P ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>(n = 4,000)</td>
<td>(n = 4,000)</td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>202</td>
<td>250</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Given that both trials used the same high-quality methodology and achieved the same level of statistical significance, some would assume we should view both results with similar confidence. But this is not the case. Although the \( P \) values in our hypothetical trials suggest the results have the same probability of representing a true finding, there is a substantial difference in the fragility of the demonstrated \( P \) values. For example, if we were to add two events to the treatment group in the first trial, the \( P \) value would become 0.13. In marked contrast, adding two events to the treatment group in the second trial results in no meaningful impact on the \( P \) value, which would remain 0.02 (table 2).

Which result in the example above is more plausible? Consider that there are at least six independent risk factors associated with perioperative myocardial infarction and that the prevalence of each of these factors in patients suffering a perioperative myocardial infarction varies from 8 to 40%. Furthermore, many of these risk factors are far more strongly associated with perioperative myocardial infarction (e.g., odds ratio 2.94 for emergent and urgent surgery) than the realistic moderate size effects we seek to identify in drug trials. It is therefore understandable how the large effect seen in our first hypothetical trial could have easily resulted from an imbalance in risk factors across the two treatment groups. In contrast, the size of our second hypothetical trial makes a meaningful imbalance in prognostic factors between the treatment groups highly unlikely. Consistent with our example, there is considerable evidence that highly cited studies in leading medical journals are frequently contradicted (16%) or shown to have reported substantially exaggerated treatment effects (16%) in subsequent studies, and the only identified factor explaining this outcome is that the initial trial had a small sample size.

Which participants and events represent the transition from a small trial to a large trial is a matter of debate and ongoing investigation. It is important to recognize that both the number of participants and events are relevant. Some data suggests that trials with event rates of 10% require, at a minimum, several thousand participants and at least 350 events and ideally 650 events to provide convincing evidence of a moderate-size treatment effect. Most RCTs require thousands of patients and several hundred events to minimize the risk of false positive or negative results.

All of this discussion relates to dichotomous outcomes (e.g., death) because patient-important outcomes are usually dichotomous. Trials evaluating continuous outcomes (e.g.,
change in blood pressure) do not require the same large sample sizes needed for trials of dichotomous outcomes. Trials that evaluate continuous outcomes do need, however, to ensure that enough patients are randomized to achieve balance of prognosis between the two treatment groups.33

Why Is It Necessary to Have Large Populations Participate In Observational Studies?

Most observational studies try to identify independent risk factors or develop a risk-prediction model. To achieve these goals researchers use statistical methods, most commonly multivariable analysis. The necessity to initially consider a host of potential risk factors creates formidable sample-size requirements. Simulation studies demonstrate that logistic models require 12–15 events per predictor to produce stable estimates.34,35 Researchers wanting to consider 10 possible predictors (a reasonable number for an initial investigation) would require 150 events; if they include a population with an event rate of 5%, this would mean that they would require a sample size of 3,000 patients.

Realistically most disease areas have a much larger pool of potential predictors. For example, in the VISION Study (ClinicalTrial.gov identifier NCT00512109), investigators are developing a model to predict major perioperative vascular complications and want to evaluate 20 preoperative patient characteristics, 25 groups of comparable surgeries with one reference low-risk surgical group, two types of surgical categories (emergent and urgent surgery) with one reference category (elective surgery), and 13 center variables. To achieve this goal, the researchers need to recruit more than 15,000 patients to ensure a stable model. To provide adequate sample sizes to allow consideration of a large number of potential independent risk factors, minimize confounding, and avoid overfitted models, observational studies require large sample sizes.36

Challenges and Limitations of Large Clinical Studies

The funding requirements of most large clinical studies are not small. These costs are, however, tiny in comparison to the healthcare costs of failing to find accurate diagnostic tests, prognostic factors, and effective preventives and treatments; it is far more costly not to undertake these studies. For example, the cost related to the complications (e.g., stroke, cancer) from hormone replacement therapy was substantially greater than the cost of the large clinical trial that evaluated hormone replacement therapy.37 Although it is easy to spend large amounts of money conducting large clinical studies, it is also important to recognize that some of the largest trials in cardiology and infectious disease were not supported by the pharmaceutical industry and cost very little.38 Given that all patients undergoing surgery will receive some form of anesthesia, it is potentially feasible to undertake large anesthetic trials at very low cost. Finally, another approach that results in substantial cost savings per intervention evaluated is to increase the use of factorial designs to evaluate more than one intervention in each clinical trial.

Large clinical studies require many centers, often in many countries. Although this is a substantial challenge, the upside of requiring a large number of centers to participate is that only studies viewed as clinically important by a large number of physicians will be undertaken. Many large clinical studies have design limitations. For example, many large clinical studies exclude important patient populations (e.g., the elderly). In general, more inclusive eligibility criteria and relying on physician’s judgment of uncertainty have greater potential to help physicians in making relevant clinical decisions than restrictive eligibility criteria.

Meta-analyses

We have advocated the need to conduct large observational studies and RCTs. The same arguments apply to meta-analyses. Many hierarchies of evidence consider a meta-analysis the highest level of evidence. Although meta-analysis is a powerful research design, there are many caveats regarding their interpretation. Meta-analyses of small trials are susceptible to publication bias. Further, their results are commonly fragile, as is the case with individual small trials, as discussed above. Therefore, a meta-analysis of small studies should be viewed as hypothesis-generating, and does not negate the need for a large definitive clinical study. A meta-analysis may provide important insights even when there is a single large clinical study and several small studies addressing the same question. The results of such meta-analyses are, however, commonly dominated by the large study. The most informative meta-analyses include several large clinical studies and allow researchers to evaluate the impact across variations in patient populations, ideally through a meta-analysis of individual patient-level data.

A Way Forward

Perioperative medicine can learn from other specialties that have developed a culture for undertaking large clinical studies. For example, there has been a substantial decrease in death rates among patients with human immunodeficiency virus and patients suffering an acute myocardial infarction.39–42 These benefits were seen after large international RCTs established effective interventions that were then subsequently integrated into clinical practice.40,42 It is worth noting that even cardiology, a specialty in which large clinical trials are now common, did not have a single large RCT 25 yr ago.

Perioperative medicine would benefit from the same culture change that occurred in cardiology a few decades ago. To transition to a discipline that undertakes rigorous and reliable large clinical studies, the perioperative medicine community needs to embrace, support, and demand large clinical trials.
studies. Further, there is a need to get more perioperative physicians doing graduate programs in clinical epidemiology, enhance collaboration across disciplines (e.g., anesthesia, medicine, and surgery), lobby granting agencies to target research funding for this major neglected public health problem, and understand that large clinical studies are needed to produce reliable results.

Although these challenges are formidable, large clinical studies in perioperative medicine are starting to happen. In the noncardiac surgery setting, several large clinical studies are underway. For example, the POISE-2 Trial is a 10,000-patient factorial trial of low-dose clonidine versus placebo and low-dose acetyl-salicylic acid versus placebo in patients undergoing noncardiac surgery. The ENIGMA-II Trial is a 7,000-patient RCT evaluating nitrous oxide-containing versus nitrous oxide-free anesthesia for noncardiac surgery. The VISION Study is a prospective cohort study evaluating a representative sample of 40,000 patients age 45 yr or greater undergoing noncardiac surgery that requires hospital admission. In the cardiac surgery setting, several large trials are also underway. The CORONARY Trial is randomizing 4,700 patients to on-pump versus off-pump coronary artery bypass surgery. The SIRS Trial is randomizing 7,500 patient to methylprednisolone versus placebo in patients undergoing cardiopulmonary bypass. The ATACAS Trial is randomizing 4,600 patients in a factorial trial of acetyl-salicylic acid versus placebo and tranexamic acid versus placebo in patients undergoing coronary artery bypass grafting surgery. These trials are a start, but there is a need to build on this moment and transition perioperative medicine into a discipline that requires and undertakes large clinical studies to inform clinical care.

Conclusions

Major perioperative complications are a major serious public health problem. Substantial improvements in health outcomes are possible, but achieving this goal will require large reliable international clinical studies. If we create a culture change of acceptance and expectation of large perioperative clinical studies and see a major increase in the involvement of physicians doing graduate programs in clinical epidemiology, enhance collaboration across disciplines (e.g., anesthesia, medicine, and surgery), lobby granting agencies to target research funding for this major neglected public health problem, and understand that large clinical studies are needed to produce reliable results.

References


15. Williams RJ, Motteram JM, Sharp CH, Gallagher PJ; Dietary vitamin E and the attenuation of early lesion development in modified Watanabe rabbits. Atherosclerosis 1992; 94:153–9


