

Lost in Translation

The 2016 John W. Severinghaus Lecture on Translational Research

Daniel I. Sessler, M.D.

IN 1954, the year I was born, Dylan Thomas wrote, “Time held me green and dying, but I sang in my chains like the sea.” In these lines, he expresses his disdain for aging, illness, infirmity, and eventual death. How differently the great poet must have felt 2 yr before his premature death when he penned his most famous lines:

Do not go gentle into that good night,
Old age should burn and rave at close of day;
Rage, rage against the dying of the light.

Consider a young mother with cancer. Consider a child with a lethal congenital condition. Rage seems the only appropriate response to the dying of the light. But of course, none of us wants to go into that good night any earlier than strictly necessary, and preferably only after long and fulfilling lives. *Neither do our patients.* After all, the one thing patients ask of us, above all else, is to keep them alive. It is thus reasonable to ask how well we do. The answer depends on which perioperative period we consider.

Few patients die during surgery, and to our credit, intraoperative mortality is now at least a factor of 10 less than it was three decades ago during my residency—despite surgical patients now being much older and sicker.¹ In fact, intraoperative mortality is now so low that it is hard to measure.² The marked reduction in intraoperative mortality did not happen by magic; it happened because of a concerted effort to improve drugs, monitors, and training. No other specialty has remotely reduced mortality by an order of magnitude, and we deserve credit for the impressive improvement.

Many anesthesiologists and surgeons incorrectly believe that a patient safely delivered to the postanesthesia care unit has survived the most dangerous part of hospitalization. In fact, 30-day postoperative mortality is 1,000 times greater than preventable intraoperative mortality. If the 30 days after surgery were considered a distinct disease, it would be the third leading cause of death (fig. 1).³ The numbers are sobering: about 2% of U.S. surgical inpatients die within 30 days.⁴ Worldwide, at least five million patients die each year within a month of surgery. Furthermore, about half

of 30-day mortality occurs during the initial hospitalization—and therefore while patients remain under full medical care and in our highest-level facilities. Because patients die after surgery rather than intraoperatively, postoperative mortality must be considered *the major perioperative problem* (table 1).

During the first postoperative year, about 5% of surgical patients die. Among those more than 65 yr of age—about one third of our patients—1-yr mortality is 10%.⁵ How many anesthesiologists appreciate that one in 10 elderly surgical patients is dead within the year? Most postoperative mortality is, naturally, consequent to severe underlying pathology and necessarily invasive operations. And as might thus be expected, postoperative deaths are nonrandom: sicker patients are far more likely to die. In fact, death can be predicted remarkably accurately just from administrative data, specifically a patient’s accumulated diagnostic and procedural codes.^{6,7}

The question, then, is the extent to which anesthesiologists contribute to mortality by what we currently do, and—more importantly perhaps—whether we can prevent serious complications and mortality by doing things differently? Perhaps the place to start is with the causes of death. Thirty-day all-cause mortality is largely cardiovascular—mostly myocardial infarctions.⁸ The incidence of postoperative myocardial infarctions is far higher than generally appreciated. About 8% of surgical inpatients more than 45 yr of age have an infarction, usually within the initial 3 postoperative days.⁴ This is orders of magnitude greater than the risk in comparable patients who do not have surgery.

The infarction incidence is higher than generally appreciated because 80% of postoperative myocardial injury is clinically silent; that is, detectable only by troponin monitoring. It is tempting to assume that clinically apparent events are the more serious ones, and that others are just “troponitis.” But that would be wrong: mortality is nearly identical for symptomatic and asymptomatic postoperative

This article is featured in “This Month in Anesthesiology,” page 1A. This article is based on the John W. Severinghaus Lecture delivered by the author at the October 25, 2016, American Society of Anesthesiologists Annual Meeting in Chicago. A video version of Dr. Sessler’s October 25, 2016, Severinghaus lecture is available at <https://vimeo.com/190762203/b37fa762d4>.

Submitted for publication October 11, 2016. Accepted for publication December 7, 2016. From the Department of Outcomes Research, Anesthesiology Institute, Cleveland Clinic, Cleveland, Ohio.

Copyright © 2017, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. Anesthesiology 2017; 126:995–1004

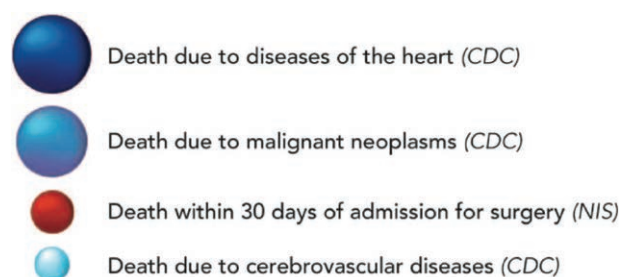


Fig. 1. Mortality within 30 days of surgery is the third leading cause of death in the United States. CDC = Centers for Disease Control; NIS = National Inpatient Sample. Reprinted with permission from Bartels K, Karhausen J, Clambey ET, Grenz A, Eltzschig HK: Perioperative organ injury. *ANESTHESIOLOGY* 2013; 119:1474–89.³

infarctions. Furthermore, the mortality is a staggering 10%. One in 10 patients with symptomatic or asymptomatic postoperative troponin elevation thus die within the month (table 2).⁹

During the initial postoperative year, the causes of death shift. About half of 1-yr mortality is due to cancer. Of course, this does not imply that surgery or anesthesia causes cancer. These are patients who come to us with malignancies and then die from disease progression. But it certainly begs the question of whether anything we do might reduce the risk of cancer recurrence. And while perhaps unlikely, there are reasons to believe that regional analgesia might help.¹⁰ At least two major trials of regional analgesia and cancer recurrence are in progress. It is also possible that perioperative administration of cyclooxygenase-2 inhibitors will reduce cancer recurrence.^{11–13} These theories remain entirely speculative, but are examples of research that would “make a difference.”

My father used to say that most any problem could be solved by “throwing money at it.” I am afraid perioperative mortality will not be so easily solved—although money would surely help! The problems are multifactorial, which is another way of saying that there is plenty of blame to pass around. Basic scientists, translational investigators, and clinicians have all contributed. In the subsequent three sections, I will identify issues each group might consider.

Table 1. Causes of 30-day and 1-yr Postoperative Mortality

Preventable intraoperative anesthetic mortality is rare
Probably about 1 per 100,000 cases in the United States
30-day all-cause mortality
≈1% in the United States
≈2% worldwide for inpatients ≥ 45 yr old
About half occurs during initial hospitalization
About one quarter is cardiovascular or consequent
1-year all-cause mortality is about 5%
Mortality is 10% in patients ≥ 65 yr old
About half from cancer

Table 2. Even Small Increases in Maximum Postoperative Serum Troponin Concentration Are Associated with Large Increases in 30-day Mortality

Peak Troponin, ng/ml	30-day Mortality, %	Time to Death, days
< 0.01	1	—
0.02	4	13
0.03–0.29	9	9
≥ 0.3	17	6

Most postoperative troponin elevation is asymptomatic, and mortality is similar in patients with and without symptoms. Data reproduced with permission from JAMA. 2012. 307(21):2295–304. Copyright © 2012 American Medical Association. All rights reserved.⁴

Basic Science

Beautiful science with no conceivable direct benefit to humans may be well worth doing. Much theoretical physics, for example, is not obviously useful but is nonetheless magnificent and broadens our understanding of the universe. There is similarly a role for fundamental mechanistic and physiologic studies. But if investigators claim that research will be useful, then it probably should be. Often it is not.

Practically every biomedical basic science grant application, and most research reports, starts with assertions that the proposed studies or presented results will markedly enhance clinical care. Few actually do. Consequently, the ratio of clinically useful advances to basic science articles is tiny. Or to put this another way, humans have proven to be a poor model for rats. Basic scientists need to help the rest of us identify studies and results that are actually applicable to patients. That is, guide us to the results that really matter and should progress to testing in animals and then humans.

We have seen many clearly delineated mechanisms that just did not translate from test tubes and animals to humans. Consider vitamins and dietary supplements; there are good mechanisms explaining why many will enhance health, yet virtually none has proven beneficial in broad populations. Vitamin E, especially, has been disappointing as large randomized trials show no benefit from supplements despite compelling reasons to anticipate benefit.^{14,15} Same with vitamin C,^{14,16,17} olive oil,¹⁸ margarine, red wine, and nearly every other dietary intervention. In fact, it is hard to think of another area where such a mountain of scientific (and nonscientific) articles have produced but a thimble-full of compelling human outcome data.

Closer to home, there is no question that nitrous oxide interferes with vitamin B12 and folate metabolism, thus increasing plasma homocysteine, impairing endothelial function, and impairing protein synthesis.¹⁹ Yet, two large randomized trials have convincingly shown that nitrous oxide causes no harm more serious than nausea and vomiting^{20,21}—and less of that than volatile anesthetics.²² Why the disconnect? Why could not our basic science colleagues help us understand that the molecular effects of nitrous oxide on protein production were unlikely to be clinically important, thus

obviating the need to randomize more than 9,000 patients to establish the safety of nitrous oxide (fig. 2)?

Therapeutic hypothermia is another example. It has been known since the early 1970s that a few degrees centigrade of hypothermia ameliorates ischemia and reperfusion injury on a cellular level.²³ Furthermore, therapeutic hypothermia reduces ischemic injury in virtually every model in every animal species.²⁴ Yet, the results in humans have been dismal. Large trials failed to demonstrate benefits from hypothermia for brain trauma,²⁵ aneurysm surgery,²⁶ and acute myocardial infarction.²⁷ (Curiously, a major trial of hypothermia for stroke, an obvious application of therapeutic hypothermia, has yet to be completed.) A bright spot was out-of-hospital cardiac arrest, based on two modest-sized studies.^{28,29} However, a subsequent study with more than twice as many

patients as the original two combined showed no benefit.³⁰ And if anything, therapeutic hypothermia for in-hospital cardiac arrest appears to worsen outcome.³¹

Even cardiopulmonary bypass, which was routinely done at 28°C for its putative brain protection, is now often conducted with patients kept normothermic with equally good results—which is consistent with many randomized trials showing no benefit from hypothermic cardiopulmonary bypass.³² At this point, neonatal asphyxia, which is reasonably well documented,^{33,34} and organ donation (based on a single major trial)³⁵ remain the only indications for deliberate hypothermia. In fairness, though, I note that hypothermia studies are challenging and that study design and execution (particularly the delay between insult and implementation of hypothermia) may be the

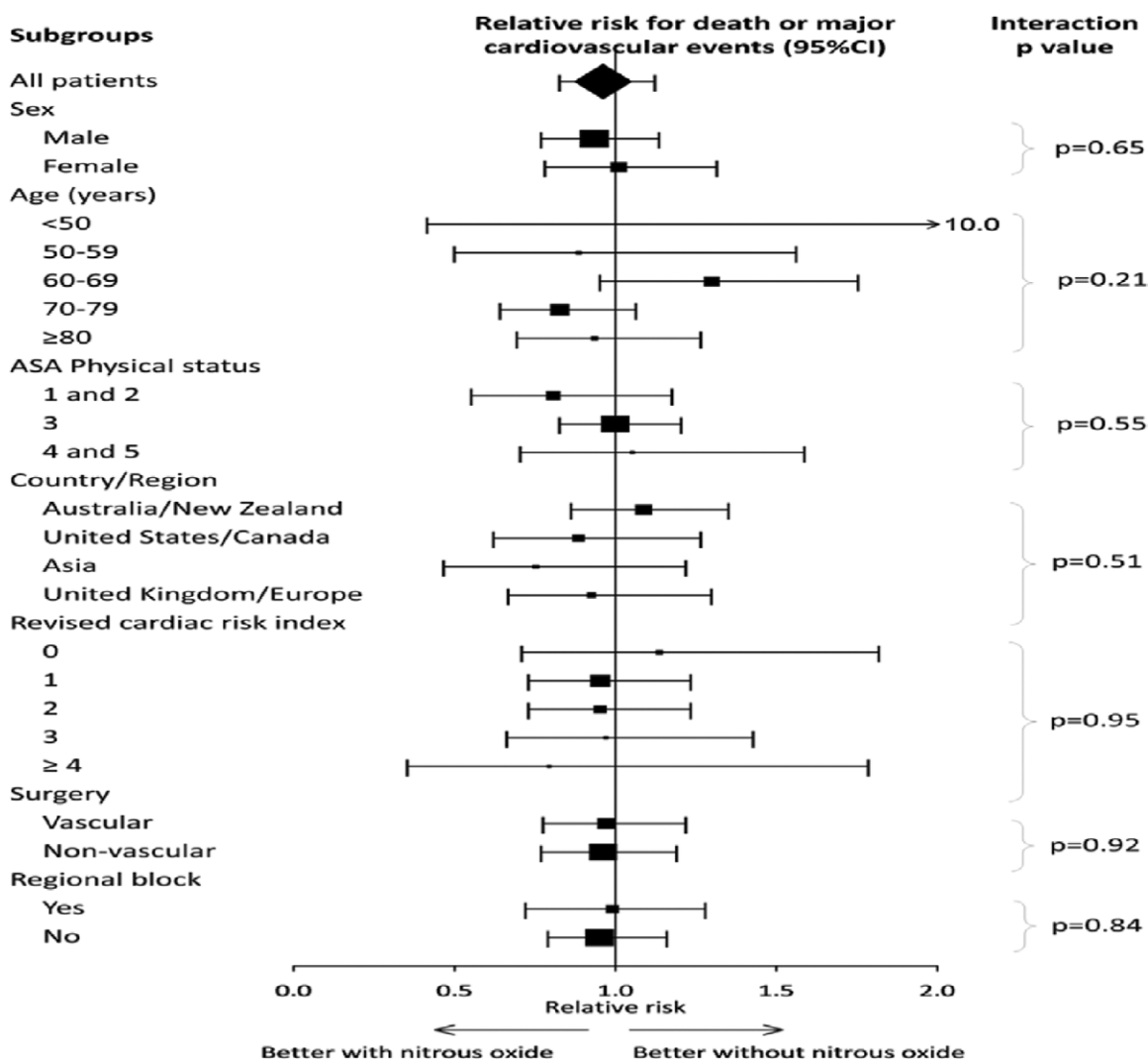


Fig. 2. In 7,112 randomized patients, 70% nitrous oxide had no adverse effect on the primary outcome of death or major cardiovascular morbidity (mostly myocardial infarction) in the entire population or in any predefined subgroups. A previous study randomized 2,050 patients to 70% nitrous oxide *versus* oxygen and also found no clear evidence of harm other than increased nausea and vomiting.²⁰ ASA = American Society of Anesthesiologists. Reprinted from The Lancet, 384, Myles PS, Leslie K, Chan MT, Forbes A, Peyton PJ, Paech MJ, Beattie WS, Sessler DI, Devereaux PJ, Silbert B, Schrickler T, Wallace S, Anzca Trials Group for the ENIGMA-II Investigators, The safety of addition of nitrous oxide to general anaesthesia in at-risk patients having major non-cardiac surgery (ENIGMA-II): A randomized, single-blind trial, 1446–54, 2014, with permission from Elsevier.²¹

major problem rather than the theory or mechanism. Half-a-dozen major trials are in progress, and therapeutic hypothermia may yet be proven beneficial in some circumstances.

A deep understanding of genetics was among the scientific triumphs of the last half-century. Powerful techniques such as genome-wide arrays were to unlock the genetic basis for much disease, opening an era of individualized medicine. While there have been undoubted advances, genetics has yet to fulfill its initial promise. Genetic analysis remains critical for diseases caused by single mutations, many of which have been understood for decades. But the more common diseases such as hypertension and cardiovascular conditions, the ones that actually kill lots of people, are controlled by dozens or hundreds of genes and have largely resisted analysis despite enormous effort.

Genetic analysis is nonetheless well on its way to replacing caffeine-halothane contracture testing for malignant hyperthermia.³⁶ Presumably, genetic analysis will eventually be the standard diagnostic approach to this uniquely anesthetic disease—and probably to many others as well. I have no doubt that genomics will eventually contribute enormously to diagnosis and treatment throughout medicine, but I am similarly impressed that progress has been much slower than predicted and anticipated.

For example, it is worth considering that National Institutes of Health (Bethesda, Maryland) spent \$15 billion dollars of its \$26 billion 2016 budget (58%) on research with key words that included “gene,” “stem cell,” and “regenerative medicine.” Perhaps as a consequence, more than 29,000 articles with those key words were published in 2014. And what do we have to show for it? Sixty years after identification of the single-gene mutation for sickle cell, not a single targeted therapy has been developed.³⁷ And sickle cell is a “simple” genetic problem. We are nowhere near solving the far more common, lethal, and complicated problems such as cardiovascular disease and cancer.

Even the most fundamental basic science is worthwhile and at least enhances understanding of physiology. Furthermore, research can be “beautiful” without being obviously useful—like astronomy. I recognize that in early stages, it can be difficult or impossible to estimate which novel techniques and approaches may prove useful. But the goal I set to scientists doing basic anesthesia research is to guide clinical investigators toward results most likely to enhance care. Or to put this another way, clinical trials are difficult, expensive, and time consuming; we will never be able to do many of them. It is thus important that we test theories that are both important and likely to be true. Basic scientists can help by guiding clinical investigators toward the theories most worth testing. New structures, such as broad-based consensus panels with various types of basic scientists and trialists, might prove helpful.

Translational and Clinical Research

There remains widespread misunderstanding about what “statistically significant” means. $P = 0.05$ does not mean that there is a 95% chance that a replication study will show similar results. Instead, $P = 0.05$ corresponds to only a 50% chance that a comparable study will have $P \leq 0.05$.^{38,39} The P value needs to be 0.005 for this replication probability to reach the conventional power criterion of 80% and 0.0003 to reach 95%.⁴⁰ Figure 3 explains the implications of $P = 0.05$ on replicability, and why a value of 0.0003 is needed to provide 95% power for replication. Typically, it requires about 3.5 times as many patients to power a study for 95% replication than for 50%. A corollary is that most clinical studies are quite underpowered for replication.

It is an unfortunate quirk of history that 0.05 was designated a “significant” P value. A more appropriate value would have been 0.005, or better 0.001.⁴¹ Had one of these values been designated the criterion for significance, medical literature would be clogged with many fewer false-positive studies—and the ones reported to be positive would be far more likely to be replicable.

A further difficulty is that “replicate” in this context applies just to the conclusion that the populations differ, not to the magnitude of the treatment effect, which is what clinicians really need to know. For example, a statistically significant result might have CIs around the relative treatment effect ranging, say, from 1.03 to 6.0. The difficulty is that a treatment effect of a few percent may not be clinically important, especially if the novel treatment is more expensive and has yet-to-be-characterized potential side effects. Conversely, a large treatment effect may be implausible and would suggest that the results are simply wrong. Large numbers of subjects are needed for robust results, especially when the outcomes of interest are relatively rare dichotomous events, such as myocardial infarction, respiratory arrest, or death (fig. 4).

Unfortunately, it takes *many* more patients to establish tight CIs around a treatment effect than to simply conclude that the populations do not differ by chance. A consequence of relying on P values as our primary strength-of-evidence indicator is that many statistically significant results have wide CIs that provide little guidance to clinicians. A further problem is that identical P values may result from studies with wildly different reliability.

For example, consider two trials of perioperative β blockers for prevention of myocardial infarctions (table 3). The first enrolls 200 patients and identifies one infarction in the treatment group and nine in the placebo group for a relative risk of 0.11 and P value of 0.02. The second enrolls 4,000 patients and identifies 200 infarctions in the treatment group and 250 in the placebo group for a relative risk of 0.80 and P value of 0.02.⁴² Which of these studies with identical P values do you believe?

The second study is far more believable for two reasons. One is that the treatment effect is plausible. That heart rate

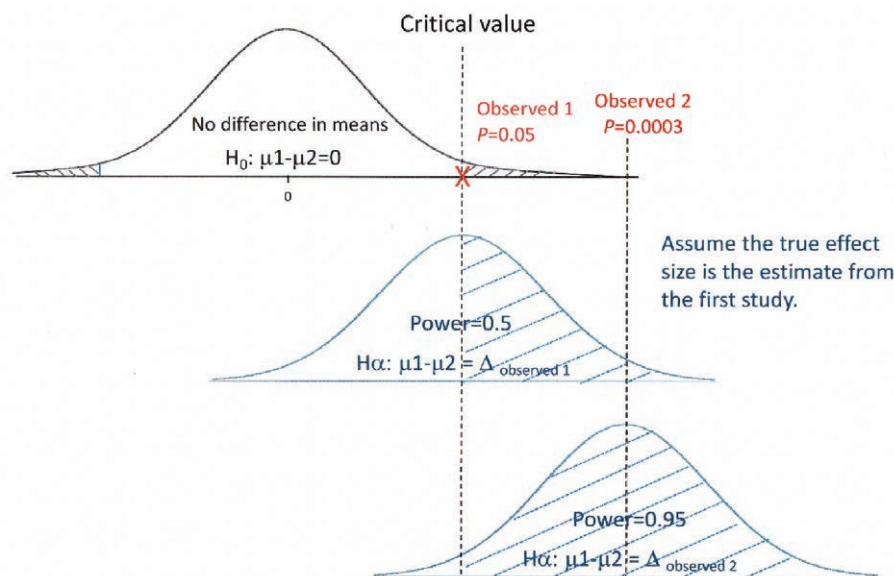


Fig. 3. “ $P = 0.05$ ” signifies that there is only 1 chance in 20 that observed distribution of results occurred by chance. However, many people believe that “ $P = 0.05$ ” means that there is a 95% chance of replicating the study. That interpretation is incorrect, and this figure demonstrates why. The *top left* figure shows the null hypothesis, which assumes no difference between two comparison groups. If the study is repeated many times, the results will cluster around zero difference in the shape of a normal distribution. For two-tailed results to be statistically significant at a $P = 0.05$, they must be 2.5% from either end of the Gaussian distribution (*i.e.*, at the *red X* in this example). We might then assume that this value is our best *estimate* of the true effect. But of course, the true effect will not exactly be this value. Instead, if the study is exactly repeated many times (with the same sample size and conditions), there will be another normal distribution centered around the initial result, which is shown in the *middle* figure. Inspection of the middle distribution clearly shows that half the values from many replications will be more extreme than the *red X* on the *top* figure and thus statistically significant at $P \leq 0.05$ (*blue shading in middle figure*). But the other half of the replication results will be to the *left* of the *red X* and thus not statistically significant. Thus, $P = 0.05$ means that the study will be replicated *only half* the time. One might then ask what initial P value is required to actually have a 95% chance of replicating a study at a $P \leq 0.05$ level. The answer is shown in the *bottom* portion of the figure and results from “sliding” the distribution around the true value to the right until 95% of it exceeds the initial observation at the *red X* (*blue shading in the bottom figure*). The *red dashed line* extending from the *center* of the *bottom* figure to the *top* figure shows that the initial P value must be less than 0.0003 to provide a 95% chance of replicating the study at a $P = 0.05$. Typically, it requires about 3.5 times as many patients to power a study for 95% replication than for 50%.

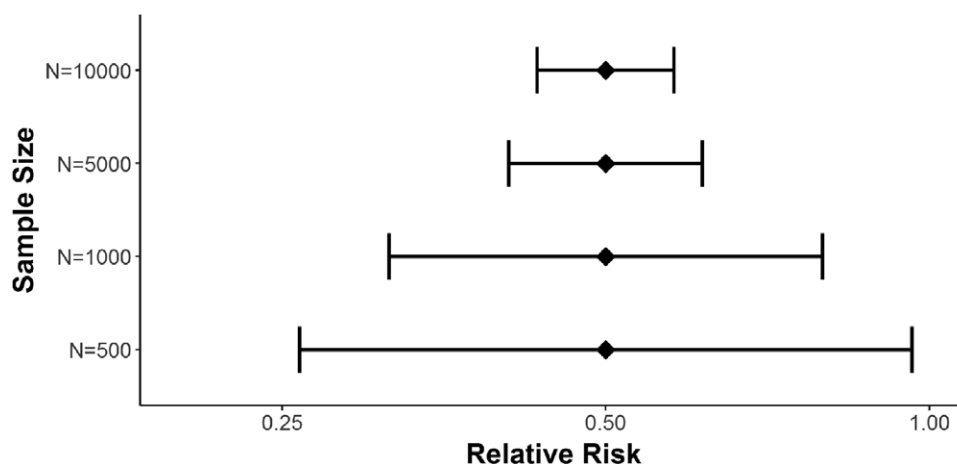


Fig. 4. An intervention reduces the risk of a major complication, say a postoperative myocardial infarction, from 10 to 5%. The relative risk with treatment is thus 0.5, and all the results shown are statistically significant. CIs around the point estimate are shown as a function of sample size. With 500 patients, the potential true value ranges over nearly a factor of 4, with the upper range nearly reaching a relative risk of 1.0 (no effect). An order of magnitude more patients are required to have reasonable confidence that the treatment effect is in fact near 50%. Large numbers of study subjects are needed to provide robust estimates of treatment effect, which is what clinicians need to guide care.

Table 3. Consider Two Identical Trials of β blockers for Prevention of Postoperative Myocardial Infarction: One with 200 Subjects and Another with 4,000

Trial	n	Treatment Infarctions	Placebo Infarctions	RR	P Value
A	200	1	9	0.11	0.02
B	4,000	200	250	0.80	0.02

One reports a 90% treatment effect and the other a 20% treatment effect; the *P* value in both cases is 0.02. Which is more plausible? The larger is more plausible for several reasons. First, it reports a reasonable relative risk (RR). That any single treatment would reduce myocardial infarctions—which are complex and multifactorial—by a factor of 10 is highly unlikely. Second, the smaller trial is fragile. Consider, for example, adding two infarctions to each treatment group, which could easily happen by chance. The smaller study would no longer be statistically significant ($P = 0.13$), whereas the *P* value in the larger trial would remain unchanged. Reprinted with permission from Devereaux PJ, Chan MT, Eisenach J, Schricker T, Sessler DI: The need for large clinical studies in perioperative medicine (editorial). *ANESTHESIOLOGY* 2012; 116:1169–75.⁴²

control reduces infarction risk by 20% is perfectly reasonable; that it would reduce the incidence of a complicated multifactorial outcome by a nearly a factor of 10 is not. The second problem is that the smaller study is fragile. The concept of fragility refers to small studies that are statistically significant, but depend critically on just a few outcomes that could easily have differed.⁴³ For example, consider the consequences of adding just two events to the treatment groups in each study, which would easily happen by chance. The *P* value for the smaller trial would increase to 0.13, but the *P* value in the larger trial would remain unchanged at 0.02. The importance of fragility is demonstrated by frequent series of progressively larger studies that “correct” initial overly optimistic results.

Most everyone is aware that random chance can falsify research results. We thus look to statistical analysis for an estimate of the extent to which apparently robust signals might result from random error (bad luck). The trouble is that there are three other major sources of error that are harder to detect and usually impossible to quantify: selection bias, confounding, and measurement bias.⁴⁴ Strong study design is the best protection against all three sources of error, with randomization generally protecting against selection bias and confounding and blinding protecting against most types of measurement bias. But even the best-designed randomized and blinded trials are subject to certain types of nonrandom error such as attrition bias.⁴⁵

Large randomized, blinded trials are generally considered the highest level of clinical evidence. But they are expensive and usually take a long time to conduct. There will never be enough randomized trials to address even a small fraction of the clinically important questions. (Novel designs such as alternating intervention⁴⁶ and decision-supported randomization⁴⁷ will help, but are only suitable for certain types of interventions.) Fortunately, trials can now be supplemented by analysis of large informative registries fed from electronic health records.⁴⁸

Registry analyses provide an opportunity to address some questions more quickly and at far lower cost than trials; furthermore, some questions such as those related to unmodifiable factors (*i.e.*, obesity, sex, and age) can only be addressed by epidemiologic analyses (table 4).³⁸ But the trade-off is that registry analyses present a far greater risk of bias and confounding without the protections of randomization and blinding. The difficulty is that few anesthesiologists—or

Table 4. Strengths and Limitations of Randomized Trials and Advantages of Registry Analyses

Randomized trials
Randomization limits selection bias and confounding
Blinding limits measurement bias
Trials are thus considered the highest level of evidence, but...
Generalizability often poor
Results date quickly as practice improves
Typically limited to two groups
Registry analyses are
Inexpensive and fast
Usually the only approach for
Nonrandomizable factors
Rare outcomes
Health services

even investigators—appreciate how subtly error can creep into noninterventional studies. Let me give you an example from a recent review.⁴⁴

Consider, for example, a study by Schull and Cobb⁴⁹ in which the investigators asked an important question: Is arthritis hereditary? The experiment consisted of asking otherwise-similar people, with and without arthritis, whether their parents had arthritis. Their results are shown in table 5. The results were clear: people with arthritis were far more likely to report that one or both parents also had arthritis. The difference was highly statistically significant, with $P = 0.003$.

There was just one problem. The subjects with arthritis and the subjects without arthritis were siblings; they had

Table 5. Otherwise-Similar People with and without Arthritis Were Asked Whether Their Parents Had Arthritis

Recall Bias, a Type of Measurement Bias		
Reported Parental Arthritis History	Rheumatoid Arthritis, %	No Rheumatoid Arthritis, %
Neither parent	27	50
One parent	58	42
Both parents	15	8

People with arthritis were far more likely to report that one or both parents also had arthritis. The difference was highly statistically significant, with $P = 0.003$. The subjects with and without arthritis, however, were siblings. They had exactly the same parents! Modified from J Chronic Dis, 22, Schull WJ, Cobb S, The intrafamilial transmission of rheumatoid arthritis. 3. The lack of support for a genetic hypothesis, 217–22, 1969, with permission from Elsevier.⁴⁹

exactly the same parents! So what happened here? Were some of the subjects lying? Unlikely. Most likely, people with rheumatoid arthritis thought much more about arthritis than those who did not. And they were far more likely to have discussed the issue with their parents and thus know (and remember) whether their parents had arthritis. This is an example of family information bias, a type of recall (measurement) bias. The difficulty is that there are many other types of bias, some of which are equally subtle, and it is usually difficult to estimate to what extent bias has degraded observational analyses. For additional discussion of sources of error and clinical research methodology, see recent reviews.^{44,45,48}

Small fragile trials and confounded registry analyses do not advance our specialty. Some even guide us in the wrong direction, producing potential or actual harm. What we need is fewer and better studies. The goal I set for clinical investigators is thus to continue the recent trend toward large well-powered studies that provide actionable answers to important questions.^{42,50}

Very large clinical trials—the ones providing the best guidance to clinicians—can require years of effort by hundreds of investigators. It is unreasonable to expect investigators to sustain such effort if they will not be rewarded with academic credit. If anesthesia is to have the number of large trials the specialty needs and deserves, department chairs and university promotion committees will have to recognize participation in large trials and consequent “corporate” authorship as a real academic activity.

Clinicians

Clinicians, you are not off the hook. Most everyone talks about practicing evidence-based medicine. And I fully understand the challenge because when you look for evidence, there turns out to be remarkably little. But it is also undoubtedly true that many clinicians do not implement well-established practices, instead basing practice on best clinical judgment—also known as “making it up.” Or even worse, there are clinicians who continue to practice much as they were taught in residency decades ago, ignoring rigorously proven advances.

For example, troponin screening for myocardial injury is inexpensive and the number needed to test is less than 15,^{9,51} which is tiny compared to many routine tests. Furthermore, troponin monitoring identifies a condition that has a stunning 10% 30-day mortality. And unlike many test results, positive troponins are actionable. Patients who experience a postoperative infarction should have a cardiology consult, have their hypertension and heart rate controlled, be put on aspirin and angiotensin-converting enzyme-inhibitors, and considered for statin treatment. Infarctions are also major life events and can be used as *teachable moments*^{52,53} to encourage patients to exercise, eat a healthful diet, and stop smoking. All these opportunities are lost in unscreened patients.⁵⁴

Troponin then is a valuable screening test that is rarely ordered. N-terminal pro-brain natriuretic peptide falls into the same category.^{55,56} But what about all the useless tests we order? How about all those coagulation tests in patients with no relevant history? What about all the electrocardiograms in asymptomatic patients, and stress tests that do not even slightly alter management?⁵⁸ Also consider perioperative normothermia for which there is copious evidence.⁵⁷ In the United States, most surgical patients are actively warmed, which is effective.⁵⁸ But the use of effective warming is much less common in other countries—and almost nonexistent in some.

For half-a-century, anesthesiologists have sniggered when internists advised us to “avoid hypoxia and hypotension.” We dismissed the advice about avoiding hypoxemia because we know it is rare intraoperatively. But we also now know that *postoperative* hypoxemia is common, severe, and prolonged—and that nurses taking conventional vital signs at 4-h intervals miss 90% of it.⁵⁹

We have not done better with hypotension: in recent years, it has become clear that intraoperative hypotension is far more common than was generally appreciated and that even mild intraoperative hypotension is strongly associated with myocardial injury and death.^{60,61} Actually, it is even worse: for decades, we fairly uncritically induced deliberate intraoperative hypotension, sometimes essentially for surgical convenience—harming who-knows-how-many patients in the process. And as with hypoxemia, it seems likely that postoperative hypotension is even more harmful. I am afraid that the internists were right, and in arrogantly dismissing their advice to “avoid hypoxia and hypotension,” we missed important opportunities to enhance care.

Nausea and vomiting prophylaxis is another area in which we generally do poorly. Too many clinicians give every patient a single antiemetic, completely ignoring highly compelling evidence and guidelines, indicating that some patients should get none—and that others should get two or more.⁶² That common and effective²² antiemetics are inexpensive is no excuse since all drugs potentially cause complications.

It is also worth noting the practices for which there was no compelling evidence. Nitrous oxide is an example I mentioned previously. Of course, nitrous oxide is hardly essential and it is perfectly easy to provide general anesthesia without the drug. But that hardly excuses allegedly scientific decisions to completely eliminate the drug in some institutions or to build new hospitals without nitrous oxide piping.

Most importantly, clinicians need to move beyond the operating room because postoperative mortality is orders of magnitude greater than intraoperative deaths. If anesthesiologists do not participate meaningfully in postoperative care, it seems unlikely that we will have any substantive impact. For example, there are already fellowships in perioperative care—offered in internal medicine departments! Hospitalists are also increasingly involved. Should not we

be training anesthesiologists in the complete medical care of postoperative patients, rather than just pain management? Postoperative care could even become a board-recognized field, joining existing subspecialties of intraoperative anesthesia, critical care, and pain management.

Similarly, it seems likely that continuous ward monitoring will soon be the standard-of-care since vital signs at 4- to 6-h intervals clearly miss many (and probably most) rescue opportunities. We should take the lead in making continuous postoperative monitoring standard and effective. And by effective monitoring, I do not mean simply purchasing devices and installed yet another computer screen in nursing stations to generate a near-constant series of false alarms that are ignored. Instead, I mean establishing integrated *systems* whereby real-time patient information, with appropriate context, streams to someone who thoughtfully evaluates data and trends and intervenes as necessary to prevent harm. Who better than an anesthesiologist?

Summary

Basic scientists, translational investigators, and clinicians all bear some responsibility for postoperative mortality, the major perioperative problem—and all can help ameliorate risk. Basic scientists, I ask you to consider which types of studies are actually likely to ultimately enhance patient care. Please also recognize that clinical research is an expensive and highly limited resource. You can help our specialty by prioritizing findings that are most likely to prove clinically important and therefore worthy of clinical investigation.

Clinical investigators need to stop churning out small fragile trials with results that are about as likely to be wrong as right—and that do not provide useful bounds on treatment effects. For registry-based studies, the problem is bias and confounding rather than fragility; neither can be “solved,” but careful question selection, study design, and analyses reduce the risk of error. What we mostly need, although, is more large randomized trials that provide robust guidance to clinicians.

And finally, clinicians need to stay abreast and implement relevant findings. It is tragic when new knowledge gained at enormous expense and effort fails to enhance care because findings are implemented slowly. But most importantly, clinicians need to “own” postoperative care rather than just managing pain. Providing good analgesia is our responsibility and a laudable goal. But pain is not the primary cause of most postoperative deaths; we need to consider and minimize *all* causes. Doing so will require that anesthesiologists substantially increase their involvement in postoperative care—that is, really become perioperative physicians.

Our specialty is at a cross-roads. One path is to embrace postoperative mortality and for basic scientists, translational investigators, and clinicians to make a sustained and concerted effort to reduce deaths after surgery—just the way our specialty solved intraoperative mortality. The other path is to declare anesthesia responsibility as ending when patients leave the postanesthesia care unit. I note, although, that the later approach is

exactly the same as defining anesthesia as irrelevant to the major perioperative problem, which is postoperative mortality.

If anesthesia is to continue making a meaningful contribution to perioperative care, we can no longer define success by getting patients to the postanesthesia care unit alive. We have largely solved intraoperative mortality, to our credit. But the operating room is no longer where patients die; instead, they die in the days and weeks after surgery. We thus need to be involved when patients actually get into trouble. Specifically, anesthesiologists need to contribute after patients leave the recovery unit. We need to actually become perioperative physicians, not just talk about it as we mostly have for the last decade or longer. This might be a good time for our specialty to remember the immortal words of Rabbi Hillel: “If I am not for myself, who will be for me? But if I am only for myself, who am I? And if not now, when?” When is now.

I recognize that perioperative medicine is a new environment for most of us. It will require new general medical and administrative skills, as well as new practice patterns. Yes, it will be hard; yes, it will require a prolonged commitment from each of us; and yes, our specialty will have to reinvent itself. When the going gets tough—as it will—I hope you will remember the words of Dory Previn:

i can't go on...
i really
can't go on;
i swear
i can't go on;
so
i guess
i'll get up
and go on.

And in doing so, we will enhance care of our patients and reinvent our specialty for the next generation. And like the great in every era, we will leave “footprints on the sands of time.”*

Research Support

Support was provided solely from institutional and/or departmental sources.

Competing Interests

The author declares no competing interests.

Correspondence

Address correspondence to Dr. Sessler: Department of Outcomes Research, Cleveland Clinic, 9500 Euclid Avenue/P77, Cleveland, Ohio 44195. DS@OR.org. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

*From A Psalm of Life, Henry Wadsworth Longfellow, 1838.

References

- Lienhart A, Auroy Y, Péquignot F, Benhamou D, Warszawski J, Bovet M, Joula E: Survey of anesthesia-related mortality in France. *ANESTHESIOLOGY* 2006; 105:1087–97
- Li G, Warner M, Lang BH, Huang L, Sun LS: Epidemiology of anesthesia-related mortality in the United States, 1999–2005. *ANESTHESIOLOGY* 2009; 110:759–65
- Bartels K, Karhausen J, Clambey ET, Grenz A, Eltzschig HK: Perioperative organ injury. *ANESTHESIOLOGY* 2013; 119:1474–89
- The Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators: Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA* 2012; 307:2295–304
- Monk TG, Saini V, Weldon BC, Sigl JC: Anesthetic management and one-year mortality after noncardiac surgery. *Anesth Analg* 2005; 100:4–10
- Sessler DI, Sigl JC, Manberg PJ, Kelley SD, Schubert A, Chamoun NG: Broadly applicable risk stratification system for predicting duration of hospitalization and mortality. *ANESTHESIOLOGY* 2010; 113:1026–37
- Dalton JE, Kurz A, Turan A, Mascha EJ, Sessler DI, Saager L: Development and validation of a risk quantification index for 30-day postoperative mortality and morbidity in noncardiac surgical patients. *ANESTHESIOLOGY* 2011; 114:1336–44
- Devereaux PJ, Sessler DI: Cardiac complications in patients undergoing major noncardiac surgery. *N Engl J Med* 2015; 373:2258–69
- The Vascular events in noncardiac Surgery patients cOhort evaluation (VISION) Investigators: Myocardial injury after noncardiac surgery: A large, international, prospective cohort study establishing diagnostic criteria, characteristics, predictors, and 30-day outcomes. *ANESTHESIOLOGY* 2014; 120: 564–78
- Sessler DI: Does regional analgesia reduce the risk of cancer recurrence? A hypothesis. *Eur J Cancer Prev* 2008; 17:269–72
- Roche-Nagle G, Connolly EM, Eng M, Bouchier-Hayes DJ, Harmey JH: Antimetastatic activity of a cyclooxygenase-2 inhibitor. *Br J Cancer* 2004; 91:359–65
- Farooqui M, Li Y, Rogers T, Poonawala T, Griffin RJ, Song CW, Gupta K: COX-2 inhibitor celecoxib prevents chronic morphine-induced promotion of angiogenesis, tumour growth, metastasis and mortality, without compromising analgesia. *Br J Cancer* 2007; 97:1523–31
- Benish M, Bartal I, Goldfarb Y, Levi B, Avraham R, Raz A, Ben-Eliyahu S: Perioperative use of beta-blockers and COX-2 inhibitors may improve immune competence and reduce the risk of tumor metastasis. *Ann Surg Oncol* 2008; 15:2042–52
- Roberts JM, Myatt L, Spong CY, Thom EA, Hauth JC, Leveno KJ, Pearson GD, Wapner RJ, Varner MW, Thorp JM Jr, Mercer BM, Peaceman AM, Ramin SM, Carpenter MW, Samuels P, Sciscione A, Harper M, Smith WJ, Saade G, Sorokin Y, Anderson GB; Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network: Vitamins C and E to prevent complications of pregnancy-associated hypertension. *N Engl J Med* 2010; 362:1282–91
- Robinson I, de Serna DG, Gutierrez A, Schade DS: Vitamin E in humans: An explanation of clinical trial failure. *Endocr Pract* 2006; 12:576–82
- Creagan ET, Moertel CG, O'Fallon JR, Schutt AJ, O'Connell MJ, Rubin J, Frytak S: Failure of high-dose vitamin C (ascorbic acid) therapy to benefit patients with advanced cancer: A controlled trial. *N Engl J Med* 1979; 301:687–90
- Greenberg ER, Baron JA, Tosteson TD, Freeman DH Jr, Beck GJ, Bond JH, Colacchio TA, Collier JA, Frankl HD, Haile RW: A clinical trial of antioxidant vitamins to prevent colorectal adenoma. Polyp Prevention Study Group. *N Engl J Med* 1994; 331:141–7
- Risk, Prevention Study Collaborative G: n-3 fatty acids in patients with multiple cardiovascular risk factors. *N Engl J Med* 2013; 368: 1800–8
- Myles PS, Chan MT, Kaye DM, McIlroy DR, Lau CW, Symons JA, Chen S: Effect of nitrous oxide anesthesia on plasma homocysteine and endothelial function. *ANESTHESIOLOGY* 2008; 109:657–63
- Myles PS, Leslie K, Chan MT, Forbes A, Paech MJ, Peyton P, Silbert BS, Pascoe E; ENIGMA Trial Group: Avoidance of nitrous oxide for patients undergoing major surgery: A randomized controlled trial. *ANESTHESIOLOGY* 2007; 107:221–31
- Myles PS, Leslie K, Chan MT, Forbes A, Peyton PJ, Paech MJ, Beattie WS, Sessler DI, Devereaux PJ, Silbert B, Schricker T, Wallace S; ANZCA Trials Group for the ENIGMA-II investigators: The safety of addition of nitrous oxide to general anesthesia in at-risk patients having major non-cardiac surgery (ENIGMA-II): A randomised, single-blind trial. *Lancet* 2014; 384:1446–54
- Apfel CC, Korttila K, Abdalla M, Kerger H, Turan A, Vedder I, Zernak C, Danner K, Jokela R, Pocock SJ, Trenkler S, Kredel M, Biedler A, Sessler DI, Roewer N; IMPACT Investigators: A factorial trial of six interventions for the prevention of postoperative nausea and vomiting. *N Engl J Med* 2004; 350:2441–51
- Mayer SA, Sessler DI: Therapeutic Hypothermia. New York, Marcel Dekker, 2004
- Dae MW, Gao DW, Sessler DI, Chair K, Stillson CA: Effect of endovascular cooling on myocardial temperature, infarct size, and cardiac output in human-sized pigs. *Am J Physiol Heart Circ Physiol* 2002; 282:H1584–91
- Clifton GL, Miller ER, Choi SC, Levin HS, McCauley S, Smith KR Jr, Muizelaar JP, Wagner FC Jr, Marion DW, Luerksen TG, Chesnut RM, Schwartz M: Lack of effect of induction of hypothermia after acute brain injury. *N Engl J Med* 2001; 344:556–63
- Todd MM, Hindman BJ, Clarke WR, Torner JC; Intraoperative Hypothermia for Aneurysm Surgery Trial (IHAST) Investigators: Mild intraoperative hypothermia during surgery for intracranial aneurysm. *N Engl J Med* 2005; 352:135–45
- Dixon SR, Whitbourn RJ, Dae MW, Grube E, Sherman W, Schaer GL, Jenkins JS, Baim DS, Gibbons RJ, Kuntz RE, Popma JJ, Nguyen TT, O'Neill WW: Induction of mild systemic hypothermia with endovascular cooling during primary percutaneous coronary intervention for acute myocardial infarction. *J Am Coll Cardiol* 2002; 40:1928–34
- Bernard SA, Gray TW, Buist MD, Jones BM, Silvester W, Gutteridge G, Smith K: Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. *N Engl J Med* 2002; 346:557–63
- Hypothermia after cardiac arrest study group: Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. *N Engl J Med* 2002; 346: 549–56
- Nielsen N, Wetterslev J, Cronberg T, Erlinge D, Gasche Y, Hassager C, Horn J, Hovdenes J, Kjaergaard J, Kuiper M, Pellis T, Stamat P, Wanscher M, Wise MP, Aneman A, Al-Subaie N, Boesgaard S, Bro-Jeppesen J, Brunetti I, Bugge JF, Hingston CD, Juffermans NP, Koopmans M, Køber L, Langørgen J, Lilja G, Møller JE, Rundgren M, Rylander C, Smid O, Werer C, Winkel P, Friberg H; TTM Trial Investigators: Targeted temperature management at 33°C versus 36°C after cardiac arrest. *N Engl J Med* 2013; 369:2197–206
- Chan PS, Berg RA, Tang Y, Curtis LH, Spertus JA; American Heart Association's Get With the Guidelines–Resuscitation Investigators: Association Between Therapeutic Hypothermia and Survival After In-Hospital Cardiac Arrest. *JAMA* 2016; 316:1375–82
- Warm Heart Investigators: Randomised trial of normothermic versus hypothermic coronary bypass surgery. *Lancet* 1994; 343: 559–63
- Azzopardi D, Strohm B, Marlow N, Brocklehurst P, Deierl A, Eddams O, Goodwin J, Halliday HL, Juszczak E, Kapellou O,

- Levene M, Linsell L, Omar O, Thoresen M, Tusor N, Whitelaw A, Edwards AD; TOBY Study Group: Effects of hypothermia for perinatal asphyxia on childhood outcomes. *N Engl J Med* 2014; 371:140–9
34. Shankaran S, Pappas A, McDonald SA, Vohr BR, Hintz SR, Yoltun K, Gustafson KE, Leach TM, Green C, Bara R, Petrie Huitema CM, Ehrenkranz RA, Tyson JE, Das A, Hammond J, Peralta-Carcelen M, Evans PW, Heyne RJ, Wilson-Costello DE, Vaucher YE, Bauer CR, Dusick AM, Adams-Chapman I, Goldstein RF, Guillet R, Papile LA, Higgins RD; Eunice Kennedy Shriver NICHD Neonatal Research Network: Childhood outcomes after hypothermia for neonatal encephalopathy. *N Engl J Med* 2012; 366:2085–92
 35. Niemann CU, Feiner J, Swain S, Bunting S, Friedman M, Crutchfield M, Broglio K, Hirose R, Roberts JP, Malinoski D: Therapeutic hypothermia in deceased organ donors and kidney-graft function. *N Engl J Med* 2015; 373:405–14
 36. Klingler W, Heiderich S, Girard T, Gravino E, Heffron JJ, Johannsen S, Jurkat-Rott K, Rüffert H, Schuster F, Snoeck M, Sorrentino V, Tegazzin V, Lehmann-Horn F: Functional and genetic characterization of clinical malignant hyperthermia crises: A multi-centre study. *Orphanet J Rare Dis* 2014; 9:8
 37. Joyner MJ, Paneth N, Ioannidis JP: What happens when underperforming big ideas in research become entrenched? *JAMA* 2016; 316:1355–6
 38. Goodman SN: Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med* 1999; 130:995–1004
 39. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG: Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *Eur J Epidemiol* 2016; 31:337–50
 40. Goodman SN: A comment on replication, p-values and evidence. *Stat Med* 1992; 11:875–9
 41. Johnson VE: Revised standards for statistical evidence. *Proc Natl Acad Sci USA* 2013; 110:19313–7
 42. Devereaux PJ, Chan MT, Eisenach J, Schricker T, Sessler DI: The need for large clinical studies in perioperative medicine. *ANESTHESIOLOGY* 2012; 116:1169–75
 43. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, Molnar AO, Dattani ND, Burke A, Guyatt G, Thabane L, Walter SD, Pogue J, Devereaux PJ: The statistical significance of randomized controlled trial results is frequently fragile: A case for a Fragility Index. *J Clin Epidemiol* 2014; 67:622–8
 44. Sessler DI, Imrey PB: Clinical research methodology 1: Study designs and methodologic sources of error. *Anesth Analg* 2015; 121:1034–42
 45. Sessler DI, Imrey PB: Clinical research methodology 3: Randomized controlled trials. *Anesth Analg* 2015; 121:1052–64
 46. Kopyeva T, Sessler DI, Weiss S, Dalton JE, Mascha EJ, Lee JH, Kiran RP, Udeh B, Kurz A: Effects of volatile anesthetic choice on hospital length-of-stay: A retrospective study and a prospective trial. *ANESTHESIOLOGY* 2013; 119:61–70
 47. Panjasawatwong K, Sessler DI, Stapelfeldt WH, Mayers DB, Mascha EJ, Yang D, Kurz A: A randomized trial of a supplemental alarm for critically low systolic blood pressure. *Anesth Analg* 2015; 121:1500–7
 48. Sessler DI, Imrey PB: Clinical research methodology 2: Observational clinical research. *Anesth Analg* 2015; 121:1043–51
 49. Schull WJ, Cobb S: The intrafamilial transmission of rheumatoid arthritis. 3. The lack of support for a genetic hypothesis. *J Chronic Dis* 1969; 22:217–22
 50. Sessler DI, Devereaux PJ: Emerging trends in clinical trial design. *Anesth Analg* 2013; 116:258–61
 51. The Postoperative Visual Loss Study Group: Risk factors associated with ischemic optic neuropathy after spinal fusion surgery. *ANESTHESIOLOGY* 2012; 116: 15–24
 52. Warner DO: Helping surgical patients quit smoking: Why, when, and how. *Anesth Analg* 2005; 101:481–7, table of contents
 53. Warner DO, Klesges RC, Dale LC, Offord KP, Schroeder DR, Shi Y, Vickers KS, Danielson DR: Clinician-delivered intervention to facilitate tobacco quitline use by surgical patients. *ANESTHESIOLOGY* 2011; 114:847–55
 54. Sessler DI, Devereaux PJ: Perioperative troponin screening. *Anesth Analg* 2016; 123:359–60
 55. Rodseth RN, Biccari BM, Chu R, Lurati Buse GA, Thabane L, Bakhai A, Bolliger D, Cagini L, Cahill TJ, Cardinale D, Chong CP, Cnotliwy M, Di Somma S, Fahrner R, Lim WK, Mahla E, Le Manach Y, Manikandan R, Pyun WB, Rajagopalan S, Radovic M, Schutt RC, Sessler DI, Suttie S, Vanniyasingam T, Waliszek M, Devereaux PJ: Postoperative B-type natriuretic peptide for prediction of major cardiac events in patients undergoing noncardiac surgery: Systematic review and individual patient meta-analysis. *ANESTHESIOLOGY* 2013; 119:270–83
 56. Rodseth RN, Biccari BM, Le Manach Y, Sessler DI, Lurati Buse GA, Thabane L, Schutt RC, Bolliger D, Cagini L, Cardinale D, Chong CP, Chu R, Cnotliwy M, Di Somma S, Fahrner R, Lim WK, Mahla E, Manikandan R, Puma F, Pyun WB, Radović M, Rajagopalan S, Suttie S, Vanniyasingam T, van Gaal WJ, Waliszek M, Devereaux PJ: The prognostic value of pre-operative and post-operative B-type natriuretic peptides in patients undergoing noncardiac surgery: B-type natriuretic peptide and N-terminal fragment of pro-B-type natriuretic peptide: A systematic review and individual patient data meta-analysis. *J Am Coll Cardiol* 2014; 63:170–80
 57. Sessler DI: Perioperative thermoregulation and heat balance. *Lancet* 2016; 387:2655–64
 58. Sun Z, Honar H, Sessler DI, Dalton JE, Yang D, Panjasawatwong K, Derose AF, Salmasi V, Saager L, Kurz A: Intraoperative core temperature patterns, transfusion requirement, and hospital duration in patients warmed with forced air. *ANESTHESIOLOGY* 2015; 122:276–85
 59. Sun Z, Sessler DI, Dalton JE, Devereaux PJ, Shahinyan A, Naylor AJ, Hutcherson MT, Finnegan PS, Tandon V, Darvish-Kazem S, Chugh S, Alzayer H, Kurz A: Postoperative hypoxemia is common and persistent: a prospective blinded observational study. *Anesth Analg* 2015; 121:709–15
 60. Mascha EJ, Yang D, Weiss S, Sessler DI: Intraoperative mean arterial pressure variability and 30-day mortality in patients having noncardiac surgery. *ANESTHESIOLOGY* 2015; 123:79–91
 61. Monk TG, Bronsert MR, Henderson WG, Mangione MP, Sum-Ping ST, Bentt DR, Nguyen JD, Richman JS, Meguid RA, Hammermeister KE: Association between intraoperative hypotension and hypertension and 30-day postoperative mortality in noncardiac surgery. *ANESTHESIOLOGY* 2015; 123:307–19
 62. Apfel CC, Kranke P, Eberhart LH, Roos A, Roewer N: Comparison of predictive models for postoperative nausea and vomiting. *Br J Anaesth* 2002; 88:234–40