SPECIAL ARTICLE

# Apophenia and anesthesia: how we sometimes change our practice prematurely

# Apophénie et anesthésie : comment il nous arrive de modifier prématurément notre pratique

Neil A. Hanson, MD · Matthew B. Lavallee, MD · Robert H. Thiele, MD

**Abstract** *Human beings are predisposed to identifying false patterns in statistical noise, a likely survival advantage during our evolutionary development. Moreover, humans seem to prefer "positive" results over "negative" ones. These two cognitive features lay a framework for premature adoption of falsely positive studies. Added to this predisposition is the tendency of journals to "overbid" for exciting or newsworthy manuscripts, incentives in both the academic and publishing industries that value change over truth and scientific rigour, and a growing dependence on complex statistical techniques that some reviewers do not understand. The purpose of this article is to describe the underlying causes of premature adoption and provide recommendations that may improve the quality of published science.*

**Résumé** *Les êtres humains ont tendance à identifier de fausses corrélations dans le bruit de fond statistique, ce qui nous a probablement conféré un avantage en matière de survie au cours de notre développement évolutionnaire. De plus, l'être humain semble préférer les résultats « positifs » aux résultats « négatifs ». Ces deux caractéristiques cognitives posent un cadre expliquant l'adoption hâtive d'études faussement positives. À cette prédisposition s'ajoutent la tendance des revues à « surenchérir » pour les manuscrits prometteurs ou notables, les incitatifs tant dans les milieux académiques qu'éditoriaux, qui préfèrent le changement à la vérité et à la rigueur scientifique, et une dépendance croissante à l'égard de techniques statistiques complexes que certains réviseurs ne comprennent pas. L'objectif de cet article est de décrire les causes sous-jacentes d'adoption prématurée de nouveautés et de proposer des recommandations afin d'améliorer la qualité de la science publiée.*

**Keywords** apophenia · anesthesia · bias · premature adoption · incentives

This article is accompanied by an editorial. Please see Can J Anesth 2021; this issue.

N. A. Hanson, MD (✉) · M. B. Lavallee, MD · R. H. Thiele, MD
Department of Anesthesiology, University of Virginia Health System, PO Box 800710, ville, VA 22908-0710, USA
e-mail: nah3q@virginia.edu

Science is in the midst of a crisis. A string of high-profile retractions and clear evidence of outright fraud, most recently related to the COVID-19 pandemic, have captured the world's attention and shaken the public's belief in scientific integrity.[1,2] Criticism of the peer review process has reached the mainstream,[3] and major news outlets now routinely report on the results of scientific trials in the pre-review stage of publication (e.g., www.medrxiv.org). This removes an important check on data integrity, through the peer review process, and allows the general public to consume "news" that has not been properly verified by subject matter experts. While intentional fabrication of data is heinous and newsworthy, it is a relatively infrequent occurrence.[4,5] A much larger, more complex, and more sinister threat to scientific data integrity is the premature acceptance of non-fraudulent data that, while scientifically valid and "statistically significant," for reasons we will describe below does not warrant wholesale adoption. It is in this space—premature adoption—that the specialty of anesthesiology (which includes critical care, perioperative

medicine, and pain management) has, like many specialties, been damaged.

The purpose of this manuscript is to describe the underlying economic, mathematical, social, and scientific causes of premature adoption. We will provide the reader with a chronological list of high-profile examples of premature adoption in three domains of anesthesiology (critical care, perioperative, and pain), and, based on both the underlying causes and notable examples in our specialty, make recommendations that may improve the quality of our own literature as well as the ability of our readership to effectively integrate anesthesiology science into their practice.

## Underlying psychology

In two contemporary works, *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*[6] and *Blink: The Power of Thinking Without Thinking*,[7] authors Nassim Taleb and Malcom Gladwell make an evolutionary argument for the premature acceptance of data. Both Taleb and Gladwell argue that millennia of external threats and selection pressures have created a species that very quickly draws conclusions and acts based on very small data sets. Let us use an example to illustrate their theory. Say, for instance, that you and your family were foraging for food in eastern Europe thousands of years ago. You stumble across *Atropa belladonna*, a tropane alkaloid-producing plant that is one of the most lethal in the Western Hemisphere. But *belladonna* also makes berries that appear enticing. Perhaps your uncle and his son both died after eating berries from *belladonna* many years ago, but you are currently starving for nourishment. What do you do? You know of only two people who have ever consumed *belladonna*, both of whom have died.

As you might imagine, humans who concluded that *Atropa belladonna* was lethal, even after only one or two observations, might be more likely to survive (and procreate). Avoiding *Atropa belladonna* is, in the circumstances described, the correct decision. Those individuals who required "more data" to be certain of the validity of their hypothesis, more than likely, perished. Clearly, this "study" is currently underpowered, and submitting a scientific study with an $n$ of 2 would be considered malpractice in 2021. But, in this specific circumstance, does it truly matter if the study is underpowered?[8] While no one would ever categorize these observations as an application of the scientific method, the experience provided "data" that changed "practice". Thousands and thousands of years of this selection process has ultimately led to a species that is primed to identify patterns that do not necessarily exist,

because for most of human history it was far safer to err on the side of *overidentification* than *underidentification* of important patterns and relationships around us.

While the *Atropa belladonna* example is an extreme example, it illustrates two key points in how humans make decisions. The first is referred to as *cost asymmetry* and it has numerous implications in religion, sociology, and economics—but also science and statistics—the sum of which are included in a relatively new field of study called *error management theory*.[9] Simply put, humans are biased towards making less costly errors even if it means increasing overall error rates. The second key point, that the human mind tends to identify patterns that do not really exist, is well documented in the psychology literature and deserves greater discussion, as it is particularly relevant to the problem of premature adoption.

Most individuals seem to develop erroneous perceptions about the meaning of random, even binary, data. Robert Ladouceur measured this directly in a series of experiments which showed that the subjects believed that objectively independent events were causally linked in some fashion.[10] This is perhaps best described in the *gambler's fallacy*— the belief, for instance, that after receiving three red numbers in a row while playing roulette, that a black number is "due".[11] The underlying thinking behind this perception has to do with how humans conceive of the concept of *chance* itself. "Chance is commonly viewed as a self-correcting process in which a deviation in one direction induces a deviation in the opposite direction to restore the equilibrium."[12] In fact, the term "corrected" should not be used to describe this process, as it is not deterministic in any fashion. Rather, we need to understand that the "deviations" are merely diluted over time as the number of events increases. This is the essence of what Amos Tversky and Daniel Kahneman referred to as the *law of large numbers*. Unfortunately, humans predominantly ascribe to the *law of small numbers*, wherein they incorrectly perceive that small samples are representative of the greater population from which they are drawn.[13]

Psychosis is simply an extreme manifestation of our propensity to create connections that do not actually exist. This understanding has led to the development of the concept of apophenia—the "tendency to perceive meaning in noise"—and "magical thinking," both of which are associated with anomalous perceptual experiences and frank psychosis.[14-17] This is of concern because, when individuals are presented with graphical depictions of synthetic data, over 60% of user-generated insights are patently false (Figure).[18] Substantial efforts have been made to identify the neuroanatomical foundation of apophenia. For instance, the use of transcranial magnetic stimulation to inhibit activity in the left lateral temporal area significantly reduces the tendency of healthy

volunteers to report meaningful information when presented with randomly generated visual noise.[19]

While the evolutionary foundation for our premature establishment of patterns and relationships that do not exist is well established, our adherence to those beliefs despite evidence to the contrary is more perplexing and deserves further analysis. The late Christopher Bernards alluded to this resistance to change in describing the term *postdural puncture headache*.[20] Despite evidence suggesting the arachnoid matter is the meningeal layer responsible for cerebrospinal fluid permeability, violation of the dura is assumed to be the cause of cerebrospinal fluid egress and the resultant headache seen following a "wet tap" during epidural placement. Bernards pointed out that just because something "made sense" to a physician did not necessarily mean that it worked.

## Positivity bias

*It is the perpetual error of the human intellect to be more moved and excited by affirmatives than by negatives.*—Francis Bacon[21]

One of the best-documented and most truth-distorting characteristics of the modern scientific publishing community is its inherent tendency to accept and publish "positive" studies. This phenomenon is institutional, more complex, and distinct from the propensity of individuals to identify false patterns described above. Sterling first quantified positive publication bias in 1959, documenting that 97% of published articles from four journals in 1955–1956 reported a rejection of the null hypothesis.[22] This observation has been validated repeatedly.[23-27]

The relatively recent advent of trial registries has made it easier to objectively test for publication bias in the biomedical literature. Simes *et al.* compared the outcomes of trials for ovarian cancer and multiple myeloma in the International Cancer Research Data Bank. They found that the published trials significantly overstated the benefit of combination therapy when compared with the pooled results of all registered trials (including trials that were not published).[28] Similar findings were reported when 487 research projects approved by the Central Oxford Research Ethics Committee were analyzed—the publication odds ratio of "positive" results was 2.28.[24] Positive publication bias has also been identified by following the trajectory of studies published in abstract form—an analysis of almost 30,000 published abstracts revealed that those reporting "positive" results were 30% more likely to be published in peer-reviewed scientific literature than those who did not.[29,30] Positive publication bias was recently identified in the anesthesiology literature, based on an analysis of 1,163 studies in 14 journals. In this analysis, positive results were associated with an increased likelihood of publication, and this effect was particularly pronounced in journals with a higher impact factor.[25] In reviewing research methodology of scientific journals explicitly dealing with the specialty of anesthesiology from 2007 to 2016, various issues with regard to trial registration and outcomes reporting were cited.[31,32] For example, in 2015, 92% of "adequately registered" trials had a discrepancy in primary or secondary endpoints favouring statistical significance.[31]

The etiology of positive publication bias is not fully understood. In an experiment involving the evaluation of medical decisions by undergraduate subjects, Jonathan Baron and John Hershey showed that evaluations of decision-makers were more positive when clinical outcomes are more favourable.[33] Mahoney *et al.* examined the influence of results on peer reviewers, by randomizing 75 referees to review one of five similar manuscripts. When reviewing nearly identical papers with either "positive" or "negative" results, referees rated both the methods and data presentation sections higher in papers that reported "positive" results, despite the fact that the methods section was identical.[34] Clearly some component of publication bias is intrinsic to human nature, and cannot be ascribed to the publication industry alone.

While not the subject of this manuscript, it is also worth pointing out that once a practice change (based on one or more "positive" studies) has been widely accepted and adopted, it can take decades to overturn even when repeated, high-quality studies indicate that the initial adoption was premature.[35] A timely example is the recent meta-analysis on the perioperative utilization of gabapentin.[36] Gabapentin had been used as early as 2004 to reduce postoperative opioid consumption in all manner of surgeries.[37] Yet, gabapentinoids were not found to improve postoperative analgesia in any significant fashion, and thus this more-than-decade-old staple of enhanced recovery programs has been shown to possibly be more harmful than helpful.

## Incentives

Physicians are incentivized to publish in order to be promoted. One has only to review any university's promotion and tenure requirements to understand that this type of scholarly work is critical to advancement. Journals also have a similar interest in publishing articles that will incite higher subscription rates. In this scenario, both actors are incentivized by the same outcome: a positive experimental result. Whether the bias is a result of authors deciding themselves not to submit negative

**Dataset 1: real (or random) choropleth maps of cancer death rates in Texas. Darker = more deaths.**
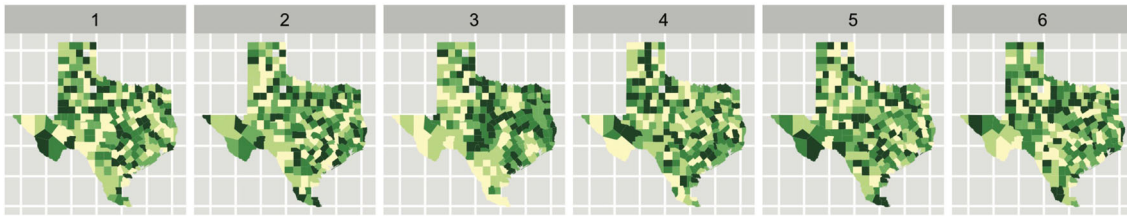


**Dataset 2: real (or random) tag cloud of words in two different book editions (larger = more words).**
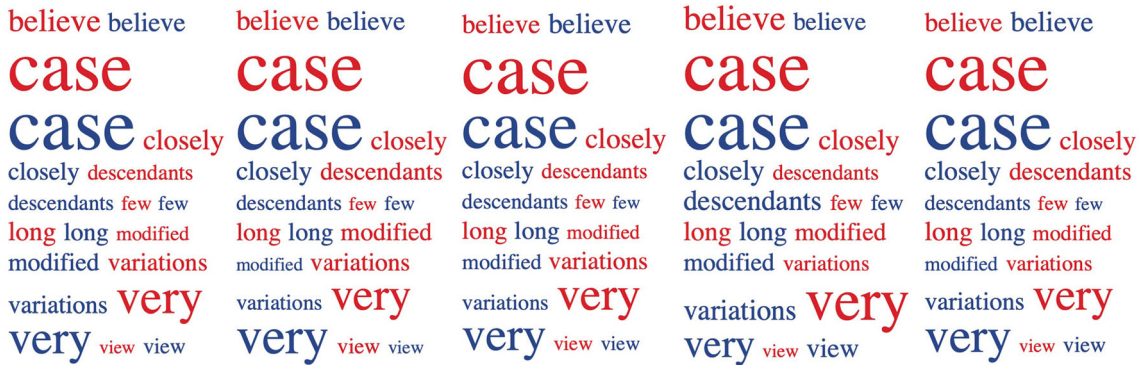


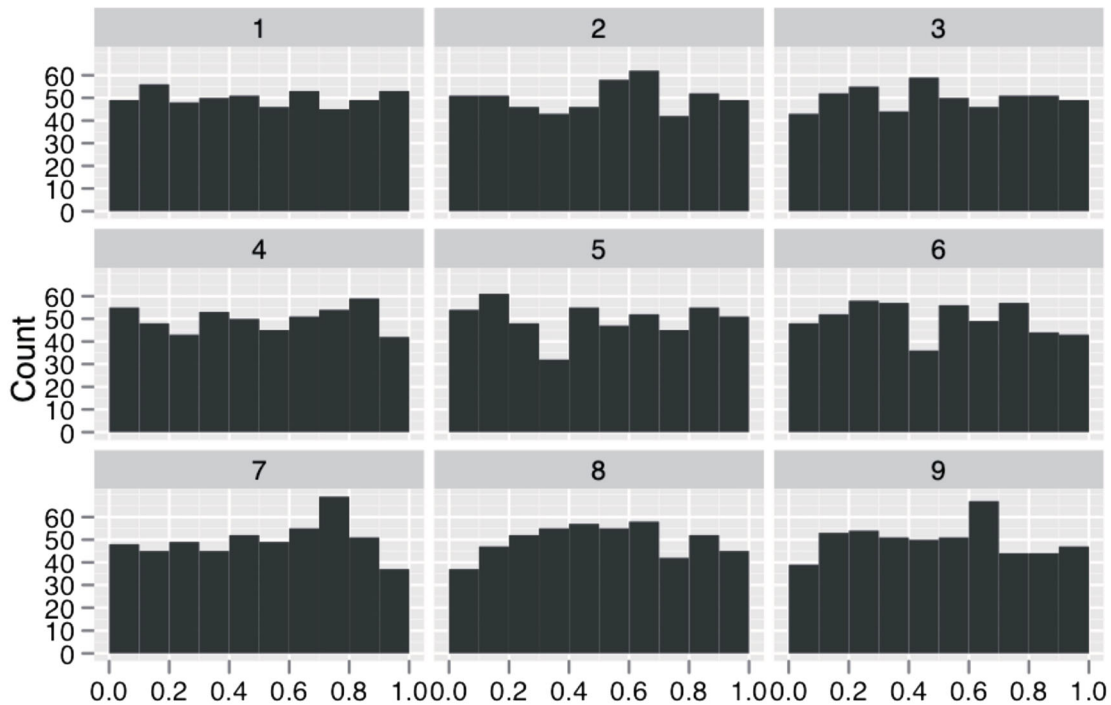**Dataset 3: real (or random) histograms of performance accuracy of nine tasks (500 participants).**



**Figure** Examples of how the ability of individuals to detect statistically meaningful relationships can be tested. Which, if any, of the above figures shows non-randomly generated data?[104] Figure A shows four randomly generated data sets and one "real" data set displaying cancer distribution in Texas—which one is "real"? Figure B displays four randomly generated word clouds, and one "real" word cloud comparing the 1st and 6th edition of Darwin's "On the Origin of Species." Which figure describes the use of words in two distinct books? Figure C depicts the distribution of performance accuracy in nine different tasks—is performance of any of these tasks distributed non-randomly, and if so, which one? (Answers: Dataset 1 = Figure 3; Dataset 2 = Figure second from right; Dataset 3 = All are randomly generated).Figure reproduced with permission from: *Wickham H, Cook D, Hofmann H, Buja A*. Graphical inference for Infovis. IEEE Trans Vis Comput Graph 2010; 16: 973-9.

results or because of the incentives journals put in place, the consequences of this synergistic relationship are the same. Interestingly, positive studies are more likely to be published in journals with higher impact factors, which has serious influence on the direction of future research.[38] Evidence of this pressure to produce "positive" results can be found by analyzing trial registries—approximately one third of randomized controlled trials publish a different primary outcome than registered in the trial.[39] Similarly, substantial discrepancies between trial registrations and published analysis are noted in 48% of published anesthesiology randomized controlled trials.[40]

Journal impact factors have become synonymous with quality.[41,42] An impact factor is determined by taking citations referencing a journal's publications divided by the number of articles published by the journal in that same time interval. Journals compete with one another to increase their impact factor, publishing more articles to which humans are subconsciously biased. While it is true that the effect this bias has on future research is undeniable, there is a more insidious repercussion: its sway on clinical practice. Because most readers are likely to choose articles from journals with higher impact factors as evidence for their medical decision-making, they will not be as exposed to articles that refute these findings.[43] Interestingly, an evaluation of the 49 most frequently cited papers addressing medical interventions from 1990 to 2004 showed that 25% of randomized controlled trials and 83% of non-randomized studies had either been exaggerated or completely contradicted.[44] Indeed, there is evidence that papers published in "high impact" journals are more likely to be retracted,[45] and that papers described in the lay press are less likely to be replicated[46] when compared with less "impactful" manuscripts.

Additional features of the publishing industry, including journal oligopoly (a small number of dominant, "high impact" journals) and artificial scarcity (deliberately restricting the number of published articles in the online era in which there are no realistic constraints to publishing space), further distorting incentives for both authors and publishers.[47]

## Statistics and faith

The inability of physicians to apply even basic statistical concepts (e.g., Bayes' theorem) in both clinical practice and in the conduct and interpretation of biomedical research is well-described.[48-51] Over the last several decades, complex statistical analysis has become an increasingly important feature of published scientific data. This presents a problem in that individual clinicians cannot independently assess the validity of a published

study, and thus have to rely on the knowledge and experience of the independent referees who manage the peer review process. As a testament to the complexity of statistical analysis, many journals now employ a statistical editor to facilitate this process. Yet, despite this, errors are still made.

In 2015, Glance et al. published a paper in Anesthesia & Analgesia in which they described the results of a retrospective analysis of 7,920 cardiac anesthetics.[52] The authors used "a fixed-effects logistic regression model that included both anesthesiologist and hospital fixed-effects" and concluded that "the rate of death or major complications among patients undergoing coronary artery bypass graft surgery varies markedly across anesthesiologists." This led to six letters to the editor, at least one of which suggested that the authors had not sufficiently shown that the distribution of mortality was different from what would be expected by random chance alone. After detailed analysis by Anesthesia & Analgesia, the manuscript was retracted—its fatal flaw was the use of fixed-effects logistic regression instead of hierarchical logistic regression as was subsequently utilized in a revised analysis.

The Glance manuscript shows the complexity and ongoing controversies around which statistical test to use when interpreting results. Most readers do not understand fixed-effects logistic regression models and therefore cannot independently verify the appropriateness of statistical techniques employed. At this level of complexity, accepting the veracity of published data has become an act of faith, because data can no longer be independently verified.

## Involvement in research

Most of the previous discussion has been centred on medicine generally, without a particular focus on anesthesiologists. Yet, the number of high-profile adoptions in the field of anesthesiology (Tables 1–3) is striking and begs the question of whether or not our specialty is for some reason predisposed to making systematic errors in interpretation and adoption of evidence-based medicine. A study by Prasad et al., which reviewed over a decade of journal articles to quantify their impact on pre-existing medical practices, found that only 38% of studies upheld existing medical practices and that 40% found evidence to the contrary.[53]

In 2006, Schwinn and Balser lamented the fact that anesthesiology departments were recipients of less than 1% of National Institutes of Health (NIH) funding from 1975 to 2003.[54] To put that into context, anesthesiologists make up almost 5% of the physician workforce in the United

**TABLE 1** Critical care

| Article topic | Year of publication | Type of study | Number of subjects | Primary outcome |
| --- | --- | --- | --- | --- |
| Methylprednisolone in the treatment of acute spinal-cord injury[78] | 1990 | RCT, placebo | 333 | Neurologic function |
| Effect of IV corticosteroids on death in acute spinal injury[79] | 2004 | RCT, placebo | 10,008 | Mortality |
| Use of PA catheters in high-risk surgical patients[80] | 1988 | RCT | 340 | Mortality |
| Cochrane review of PA catheters[81] | 2013 | Meta-analysis | 5,686 | Mortality |
| Effect of albumin in patients with cirrhosis[82] | 1999 | RCT | 126 | Mortality |
| Albumin vs saline in the ICU[83] | 2004 | RCT | 6,997 | Mortality |
| Renal improvement with dopamine[84] | 1982 | Cohort | 15 | Renal function |
| Use of dopamine in renal failure[85] | 2001 | Meta-analysis | 854 | Renal function |
| Insulin therapy in critically ill[86] | 2001 | RCT | 1,589 | Mortality |
| Intensive vs conventional insulin therapy[87] | 2009 | RCT | 6,104 | Mortality |
| Early goal-directed therapy for sepsis[88] | 2001 | RCT | 263 | Mortality |
| Early goal-directed therapy review[89] | 2015 | Meta-analysis | 4,735 | Mortality |

Important clinical trials that were controversial and in conflict with accepted practice. Unshaded studies tended to reject the null hypothesis for the primary outcome, while shaded studies did not.

ICU = intensive care unit; IV = intravenous; PA = pulmonary artery; RCT = randomized controlled trial.

**TABLE 2** Perioperative medicine

| Article topic | Year of publication | Type of study | Number of subjects | Primary outcome |
| --- | --- | --- | --- | --- |
| Effect of atenolol on mortality in non-cardiac surgery[90] | 1996 | RCT, placebo | 200 | Mortality |
| Effect of metoprolol in non-cardiac surgery[91] | 2008 | RCT, placebo | 9,298 | Mortality |
| BIS monitoring to prevent awareness[92] | 2004 | RCT, placebo | 2,463 | Awareness |
| Anesthesia awareness and the BIS[93] | 2008 | RCT, placebo | 1,941 | Awareness |
| Mortality increased in patients having "triple low"[94] | 2012 | Retrospective | 24,120 | Mortality |
| "Triple-low" alerts do not reduce mortality[95] | 2019 | RCT | 7,569 | Mortality |

Important clinical trials that were controversial and in conflict with accepted practice. Unshaded studies tended to reject the null hypothesis for the primary outcome, while shaded studies did not.

BIS = bispectral index; RCT = randomized controlled trial.

States.[55] If we are going to combat the misunderstandings that have created premature adoption, participation in the process of scientific discovery is essential. Thirteen years later, our specialty's percentage of NIH funding remains stubbornly low, at 0.6%.[56] Worse, more than half of that funding is concentrated in only ten departments, creating a winner-takes-all scenario where most academic anesthesiology departments have virtually no access to NIH funding. Contrast our specialty with medicine and surgery—as of 2020, anesthesiology departments held 508 NIH grants compared with 8,194 for internal medicine departments and 984 for surgery departments. Looking at the NIH Research Project Grant Program (R01; four to five years duration and several million USD) grants specifically, anesthesiologists held 291, internists and adult subspecialists held 3,371, and surgeons held 518. Our pipeline is at risk, as there were only 56 NIH career development awards (K-08 [basic science] and K-23 [clinical/translational]) awarded to anesthesiology departments compared with 1,022 for medicine departments and 94 for surgery departments.

Residency training programs seem to be a viable option to cultivate an interest in research, while dispelling fears and stress associated with statistics. Yet anesthesiology, as a specialty, seems to have trouble engaging its residents in research. As late as 2013, only one third of anesthesiology residency programs had a structured anesthesia residency research curriculum.[57] For the specialty of anesthesiology

**TABLE 3** Pain management

| Article topic | Year of publication | Type of study | Number of subjects | Primary outcome |
|---|---|---|---|---|
| Gabapentin improves postoperative pain[37] | 2004 | RCT | 71 | Opioid consumption |
| Perioperative use of gabapentin for acute pain[36] | 2020 | Meta-analysis | 24,682 | Pain |
| Liposomal bupivacaine in bunionectomy[96] | 2011 | RCT, placebo | 193 | Pain |
| Liposomal bupivacaine review[97] | 2021 | Meta-analysis | 619 | Pain |
| Pectoral I + II nerve block for breast cancer surgery[98] | 2015 | RCT | 120 | Pain |
| Pectoral nerve block I for breast cancer surgery[99] | 2018 | RCT, placebo | 120 | Pain |
| Addition of IPACK to AC block reduces pain after TKA[100] | 2019 | RCT | 86 | Pain with ambulation |
| The effect of IPACK block on pain after TKA[101] | 2020 | RCT, placebo | 72 | Opioid consumption |
| Effect of perineural dexamethasone on duration of interscalene nerve block[102] | 2011 | RCT | 218 | Block duration |
| Perineural versus IV dexamethasone for peripheral nerve blocks[103] | 2017 | Meta-analysis | 1,076 | Block duration |

Important clinical trials that were controversial and in conflict with accepted practice. Unshaded studies tended to reject the null hypothesis for the primary outcome, while shaded studies did not.

AC = adductor canal; IPACK = Infiltration between the popliteal artery and the capsule of the posterior knee; IV = intravenous; RCT = randomized controlled trial; TKA = total knee arthroplasty.

to continue to produce meaningful research, we must begin to offer our most junior colleagues a pathway to become involved in research early in their careers.

### Rational solutions

Despite our inherent psychological biases, our limitations in understanding basic statistical concepts, and the perverse incentives that govern the selection process of scholarly work, there are steps we can take to change the direction of academic medicine. First, while the almost universal requirement of trial registration for human studies has made it easier for reviewers to confirm that a priori primary outcomes were adhered to by investigators, very few contain a detail on the planned statistical analysis. This resulting freedom could result in analysis manipulation, which could alter the outcome of the results.

Second, we should consider transforming the way in which scientific manuscripts are published. Instead of initiating the submission process at the conclusion of the experiment, some journals could make acceptance decisions for investigator-initiated studies based on the importance of the scientific question asked and the validity of the methodology alone. After all, if the question is important and the methodology is sound, the results deserve dissemination. This would be especially important for randomized controlled trials[58] and has the added benefit of giving investigators access to rigorous critique even before initiation of the trial. By drawing the focus on the process, rather than the outcome, both positive and negative results would become useful. This has been a failing for much of contemporary clinical research, the majority of which do not produce meaningful results because of the nature of incentives.[59] A transitional option that retains some of the benefits of this system is publishing the trial design and statistical analysis in journals like *Clinical Trials*.

Third, while challenging, there should be some investment in the support of multicentre collaboration. Single-institution studies, while more affordable in aggregate (although not on a per-patient basis), may not have sufficient numbers to produce the type of power necessary to detect meaningful differences, nor do the findings result in greater generalizability. The failure of the general scientific community to address inadequate statistical power is well documented.[60,61] Unfortunately, this harkens back to the incentives created by professional

advancement, which ultimately hinder collaboration between institutions. State and national societies should attempt to foster multicentre collaboration through the appropriation of grants and priority of the funded research in their journals.

Fourth, the anesthesiology community should be wary of studies whose methodology depends on statistical techniques that neither reviewers nor readers understand. When the complexity of statistical analysis is out of the ordinary or even novel, an editorial by the statistical editor of the journal can be both reassuring and educational to the scientific community at large. Collaboration between clinicians and statisticians is essential through the development of scientific trials and their final analysis.

Fifth, given our predisposition to accept false positives as well as the myriad of confounding forces in both academics and the publishing industry, the anesthesiology community should engage in serious discussion about the "one size fits all" *P* value of 0.05. Some authors have proposed simply lowering the *P* value to 0.005.[62,63] While this will surely reduce the probability of falsely rejecting the null hypothesis, this approach has the disadvantage of increasing the probability of missing a clinically meaningful and statistically significant relationship, leading to controversy.[64] One advantage of changing the *P* value to 0.005 is that it will encourage investigators to increase the sample size of most studies, which would *reduce* the probability of incorrectly accepting the null hypothesis.

The reality is that the *P* value should consider the "cost" of making a statistical error. For instance, prior to advocating wide adoption of an expensive pharmacotherapeutic agent that has a formidable side effect profile, one needs to be sure that this is justified. Other questions, such as whether or not to fill an endotracheal tube cuff with air, saline, or dilute lidocaine to prevent coughing (inexpensive and safe) may not require such rigour. Some authors have suggested that we eliminate *P* values altogether.[65] The addition of a 95% confidence interval to an effect magnitude can help put the meaning of an effect size into context and be particularly helpful for clinicians. Lastly, an increased understanding of Bayesian statistical concepts, which incorporate pre-test probabilities into outcome estimates, would be beneficial for all practicing physicians.[66]

Finally, formal education on statistics and evaluation of literature should be incorporated into anesthesiology training programs. It is fundamental that the next generation of anesthesiologists be able to critically analyze all manner of research and understand the statistical tests used by the investigators. Graduate medical education seems to have lagged behind in preparing residents for this task.[67] Though the Accreditation Council for Graduate Medical Education requires residents to demonstrate competency in statistics, in anesthesiology residency programs the topic seems to make only a token appearance around the time of formal examinations. The exposure of statistics to residents on a more constant basis provides the chance to make these adult learners more willing to venture into research. It would be unfortunate if younger physicians with a zest for research were deterred because of an aversion to statistical analysis.

Taken together, these interventions could significantly reduce the probability that anesthesiologists will prematurely adopt practice changes that later turn out to be non-beneficial or even harmful.

## Counterarguments and balance

While this manuscript has been critical of the anesthesiology community's willingness to prematurely adopt practice changes (some of which have caused harm) based on inadequate evidence, we must balance that observation with some practical realities. Large scale, multicentre, prospective randomized controlled trials are expensive and time consuming, and it is not realistic to expect that all scientific questions will be answered in this fashion. It is interesting to note that anesthesia mortality has decreased *dramatically* over the last 80 years despite a dearth of large scale, multicentre randomized controlled trials targeted to anesthesiology specifically.[68] Much of this improvement is thought to be due to technological advances.[69] Nevertheless, the landmark pulse oximeter trial including over 20,000 participants did not find a difference in perioperative mortality.[70] Similarly, while capnography is regarded by the World Health Organization as an essential intraoperative monitoring device, there are no large-scale prospective randomized controlled trials of this device in the intraoperative environment, and most evidence to support this technology comes either from large retrospective analyses (including closed claims databases) or small prospective trials with conflicting results.[71] One can only conclude that highest-level evidence is not the only means by which patient care can advance.

While we have described a series of experimental trials which were, in many cases, adopted prematurely, it is important to also acknowledge that many high-quality studies have been performed, the findings adopted, and not overturned. Additionally, experimental studies, even small ones, may help overturn logical but incorrect assumptions made by clinicians. For instance, while it is logical that cardiopulmonary resuscitation (CPR) with both chest compressions and ventilation would be superior to

compression-only CPR, in the out-of-hospital environment compression-only CPR appears to be superior.[72] Our point is not that these smaller trials should not be conducted, but that they should be interpreted with caution.

Observational data also has its role—it is easier to perform, does not involve the ethical dilemma of an intervention, and is especially useful for hypothesis generation. Interestingly, some studies have suggested that observational trials produce similar effect estimate differences compared with prospective randomized controlled trials,[73-75] although not all analyses agree.[76,77]

Lastly, we strongly believe that science is never "settled." Some of the large experimental trials cited in this manuscript as being "definitive" will undoubtedly be overturned in the future. It is our duty as clinicians and investigators to always view scientific evidence with an appropriate level of caution and humility, while at the same time not becoming overly agnostic and refusing to adopt practice changes because there isn't any strong evidence to support a change. For certain clinical questions, studies may not yet exist, and the clinician must act based on his or her own knowledge of physiology, pharmacology, aggregation of smaller, underpowered studies, or, in extreme cases, extrapolation of preclinical data.

## Conclusion

Human beings are predisposed to identifying false patterns in statistical noise—this was likely a survival advantage during our evolutionary development, and has been shown repeatedly in modern humans using a variety of neurocognitive tests. In addition, humans seem to prefer "positive" results over "negative" ones. These two cognitive features lay a framework for premature adoption of falsely positive ideas. Added to this predisposition is the tendency of journals to "overbid" for exciting or newsworthy manuscripts, incentives in both the academic and publishing industries that value change over truth and scientific rigour, and a growing dependence on complex statistical techniques that some reviewers do not understand. These features may partially explain why the anesthesiology community has repeatedly adopted practice changes based on small, spuriously positive studies that were later overturned. We suggest improvements in the scientific publication process, increasing incentives for multi-centred studies, decreased reliance on complex statistics, lowering the acceptable $P$ value, and targeted education on both statistics and evaluation of scientific literature. Taken together, these steps may substantially improve the quality of published science while reducing the premature adoption of falsely published studies within the anesthesiology community.

## References

1. *Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN*. Cardiovascular disease, drug therapy, and mortality in Covid-19. N Engl J Med 2020; 382: e102.
2. *Mehra MR, Desai SS, Ruschitzka F, Patel AN*. RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet 2020. https://doi.org/10.1016/s0140-6736(20)31324-6.
3. *Campbell H*. The Corruption of Peer Review Is Harming Scientific Credibility. Wall Street Journal; 13 July 2014. Available from URL: https://www.wsj.com/articles/hank-campbell-the-corruption-of-peer-review-is-harming-scientific-credibility-1405290747 (accessed February 2021).
4. *Fanelli D*. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One 2009; DOI: https://doi.org/10.1371/journal.pone.0005738.
5. *George SL, Buyse M*. Data fraud in clinical trials. Clin Investig (Lond) 2015; 5: 161-73.
6. *Taleb NN*. Fooled By Randomness : the Hidden Role of Chance Life and in the Markets. NY: Texere; 2001.
7. *Gladwell M*. Blink: The Power Of Thinking Without Thinking. NY: Back Bay Books, Little, Brown And Co.; 2005.
8. *Yeh RW, Valsdottir LR, Yeh MW, et al*. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. BMJ 2018; DOI: https://doi.org/10.1136/bmj.k5094.
9. *Haselton MG, Buss DM*. Error management theory: a new perspective on biases in cross-sex mind reading. J Pers Soc Psychol 2000; 78: 81-91.
10. *Ladouceur R, Paquet C, Dubé D*. Erroneous perceptions in generating sequences of random events. J Appl Soc Psychol 1996; 26: 2157-66.
11. *Sundali J, Croson RT*. Biases in casino betting: the hot hand and the gambler's fallacy. Judgm Decis Mak 2006; 1: 1-12.
12. *Tversky A, Kahneman D*. Judgment under uncertainty: heuristics and biases. Science 1974; 185: 1124-31.
13. *Tversky A, Kahneman D*. Belief in the law of small numbers. Psychol Bull 1971; 76: 105-10.
14. *Fyfe S, Williams C, Mason OJ, Pickup GJ*. Apophenia, theory of mind and schizotypy: perceiving meaning and intentionality in randomness. Cortex 2008; 44: 1316-25.
15. *van Elk M*. Paranormal believers are more prone to illusory agency detection than skeptics. Conscious Cogn 2013; 22: 1041-6.
16. *Blain SD, Longenecker JM, Grazioplene RG, Klimes-Dougan B, DeYoung CG*. Apophenia as the disposition to false positives: a

unifying framework for openness and psychoticism. J Abnorm Psychol 2020; 129: 279-92.

17. *Shafer SL, Dexter F*. Publication bias, retrospective bias, and reproducibility of significant results in observational studies. Anesth Analg 2012; 114: 931-2.

18. *Zgraggen E, Zhao Z, Zeleznik R, Kraska T*. Investigating the effect of the multiple comparisons problem in visual analysis. Proceedings of the Conference on Human Factors in Computing Systems 2018: 1-12.

19. *Bell V, Reddy V, Halligan P, Kirov G, Ellis H*. Relative suppression of magical thinking: a transcranial magnetic stimulation study. Cortex 2007; 43: 551-7.

20. *Bernards CM*. Sophistry in medicine: lessons from the epidural space. Reg Anesth Pain Med 2005; 30: 56-66.

21. *Donaldson I*. Francis Bacon's comments on the power of negative observations in his Novum Organum, first published in 1620. J R Soc Med 2016; 109: 459-60.

22. *Sterling TD*. Publication decisions and their possible effects on interference drawn from tests of significance or vice versa. J Am Stat Assoc 1959; 54: 30-4.

23. *Dickersin K*. The existence of publication bias and risk factors for its occurrence. JAMA 1990; 263: 1385-9.

24. *Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR*. Publication bias in clinical research. Lancet 1991; 337: 867-72.

25. *De Oliveira GS, Jr Chang R, Kendall MC, Fitzgerald PC, McCarthy RJ*. Publication bias in the anesthesiology literature. Anesth Analg 2012; 114: 1042-8.

26. *Nissen SB, Magidson T, Gross K, Bergstrom CT*. Publication bias and the canonization of false facts. eLife 2016; https://doi.org/10.7554/elife.21451.

27. *Fanelli D, Costas R, Ioannidis JP*. Meta-assessment of bias in science. Proc Natl Acad Sci USA 2017; 114: 3714-9.

28. *Simes RJ*. Publication bias: the case for an international registry of clinical trials. J Clin Oncol 1986; 4: 1529-41.

29. *Scherer RW, Langenberg P, von Elm E*. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev 2007; DOI: https://doi.org/10.1002/14651858.MR000005.pub3.

30. *Scherer RW, Meerpohl JJ, Pfeifer N, Schmucker C, Schwarzer G, von Elm E*. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev 2018; DOI: https://doi.org/10.1002/14651858.MR000005.pub4.

31. *Jones PM, Chow JT, Arango MF, et al*. Comparison of registered and reported outcomes in randomized clinical trials published in anesthesiology journals. Anesth Analg 2017; 125: 1292-300.

32. *Chow JT, Turkstra TP, Yim E, Jones PM*. Sample size calculations for randomized clinical trials published in anesthesiology journals: a comparison of 2010 versus 2016. Can J Anesth 2018; 65: 611-8.

33. *Baron J, Hershey JC*. Outcome bias in decision evaluation. J Pers Soc Psychol 1988; 54: 569-79.

34. *Mahoney MJ*. Publication prejudices: an experimental study of confirmatory bias in the peer review system. Cogn Ther Res 1971; 1: 161-75.

35. *Tatsioni A, Bonitsis NG, Ioannidis JP*. Persistence of contradicted claims in the literature. JAMA 2007; 298: 2517-26.

36. *Verret M, Lauzier F, Zarychanski R, et al*. Perioperative use of gabapentinoids for the management of postoperative acute pain: a systematic review and meta-analysis. Anesthesiology 2020; 133: 265-79.

37. *Dierking G, Duedahl TH, Rasmussen ML, et al*. Effects of gabapentin on postoperative morphine consumption and pain after abdominal hysterectomy: a randomized, double-blind trial. Acta Anaesthesiol Scand 2004; 48: 322-7.

38. *Frank E*. Authors' criteria for selecting journals. JAMA 1994; 272: 163-4.

39. *Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P*. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009; 302: 977-84.

40. *De Oliveira GS, Jr Jung MJ, McCarthy RJ*. Discrepancies between randomized controlled trial registry entries and content of corresponding manuscripts reported in anesthesiology journals. Anesth Analg 2015; 121: 1030-3.

41. *Brown H*. How impact factors changed medical publishing–and science. BMJ 2007; 334: 561-4.

42. *McVeigh ME, Mann SJ*. The journal impact factor denominator: defining citable (counted) items. JAMA 2009; 302: 1107-9.

43. *Callaham M, Wears RL, Weber E*. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA 2002; 287: 2847-50.

44. *Ioannidis JP*. Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005; 294: 218-28.

45. *Fang FC, Casadevall A*. Retracted science and the retraction index. Infect Immun 2011; 79: 3855-9.

46. *Dumas-Mallet E, Smith A, Boraud T, Gonon F*. Poor replication validity of biomedical association studies reported by newspapers. PLoS One 2017; DOI: https://doi.org/10.1371/journal.pone.0172650.

47. *Young NS, Ioannidis JP, Al-Ubaydli O*. Why current publication practices may distort science. PLoS Med 2008; DOI: https://doi.org/10.1371/journal.pmed.0050201.

48. *Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S*. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest 2007; 8: 53-96.

49. *Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G*. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. Ann Intern Med 2012; 156: 340-9.

50. *Anderson BL, Williams S, Schulkin J*. Statistical literacy of obstetrics-gynecology residents. J Grad Med Educ 2013; 5: 272-5.

51. *Johnston BC, Alonso-Coello P, Friedrich JO, et al*. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. CMAJ 2016; 188: 25-32.

52. *Glance LG, Kellermann AL, Hannan EL, et al*. The impact of anesthesiologists on coronary artery bypass graft surgery outcomes. Anesth Analg 2015; 120: 526-33.

53. *Prasad V, Vandross A, Toomey C, et al*. A decade of reversal: an analysis of 146 contradicted medical practices. Mayo Clin Proc 2013; 88: 790-8.

54. *Schwinn DA, Balser JR*. Anesthesiology physician scientists in academic medicine: a wake-up call. Anesthesiology 2006; 104: 170-8.

55. *Association of American Medical Colleges*. Active Physicians by Sex and Specialty, 2015. Available from URL: https://www.aamc.org/data-reports/workforce/interactive-data/active-physicians-sex-and-specialty-2015 (accessed February 2021).

56. *Chandrakantan A, Adler AC, Stayer S, Roth S*. National Institutes of Health-funded anesthesiology research and anesthesiology physician-scientists: trends, promises, and concerns. Anesth Analg 2019; 129: 1761-6.

57. *Ahmad S, De Oliveira G, Jr McCarthy RJ*. Status of anesthesiology resident research education in the United States: structured education programs increase resident research productivity. Anesth Analg 2013; 116: 205-10.

58. *Law LS, Lo EA*. A two-stage review process for randomized controlled trials: the ultimate solution for publication bias? Can J Anesth 2016; 63: 1381-2.

59. *Ioannidis JP*. Why most clinical research is not useful. PLoS Med 2016; DOI: https://doi.org/10.1371/journal.pmed.1002049.

60. *Maxwell SE*. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. Psychol Methods 2004; 9: 147-63.

61. *Ahmed Ali U, Ten Hove JR, Reiber BM, van der Sluis PC, Besselink MG*. Sample size of surgical randomized controlled trials: a lack of improvement over time. J Surg Res 2018; 228: 1-7.

62. *Ioannidis JP*. The proposal to lower P value thresholds to .005. JAMA 2018; 319: 1429-30.

63. *Benjamin DJ, Berger JO, Johannesson M, et al*. Redefine statistical significance. Nat. Hum Behav 2018; 2: 6-10.

64. *de Ruiter J*. Redefine or justify? Comments on the alpha debate. Psychon Bull Rev 2019; 26: 430-3.

65. *Amrhein V, Greenland S, McShane B*. Scientists rise up against statistical significance. Nature 2019; 567: 305-7.

66. *Houle TT, Turner DP*. Bayesian statistical inference in anesthesiology. Anesthesiology 2013; 119: 4-6.

67. *Windish DM, Huot SJ, Green ML*. Medicine residents' understanding of the biostatistics and results in the medical literature. JAMA 2007; 298: 1010-22.

68. *Bainbridge D, Martin J, Arango M, Cheng D, Evidence-based Peri-operative Clinical Outcomes Research (EPiCOR)Group*. Perioperative and anaesthetic-related mortality in developed and developing countries: a systematic review and meta-analysis. Lancet 2012; 380: 1075-81.

69. *Institute of Medicine (US) Committee on Quality of Health Care in America; Kohn LT, Corrigan JM, Donaldson MS*. To Err is Human: Building a Safer Health System. Washington DC: National Academies Press; 2000.

70. *Moller JT, Pedersen T, Rasmussen LS, et al*. Randomized evaluation of pulse oximetry in 20,802 patients: I. Design, demography, pulse oximetry failure rate, and overall complication rate. Anesthesiology 1993; 78: 436-44.

71. *Wollner E, Nourian MM, Booth W, et al*. Impact of capnography on patient safety in high- and low-income settings: a scoping review. Br J Anaesth 2020; 125: e88-103.

72. *Hupfl M, Selig HF, Nagele P*. Chest-compression-only versus standard cardiopulmonary resuscitation: a meta-analysis. Lancet 2010; 376: 1552-7.

73. *Anglemyer A, Horvath HT, Bero L*. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev 2014; DOI: https://doi.org/10.1002/14651858.MR000034.pub2.

74. *Benson K, Hartz AJ*. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000; 342: 1878-86.

75. *Concato J, Shah N, Horwitz RI*. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med 2000; 342: 1887-92.

76. *Chalmers TC, Matta RJ, Smith H Jr, Kunzler AM*. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. N Engl J Med 1977; 297: 1091-6.

77. *Sacks H, Chalmers TC, Smith H Jr*. Randomized versus historical controls for clinical trials. Am J Med 1982; 72: 233-40.

78. *Bracken MB, Shepard MJ, Collins WF, et al*. A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal-cord injury. Results of the Second National Acute Spinal Cord Injury Study. N Engl J Med 1990; 322: 1405-11.

79. *Roberts I, Yates D, Sandercock P, et al*. Effect of intravenous corticosteroids on death within 14 days in 10008 adults with clinically significant head injury (MRC CRASH trial): randomised placebo-controlled trial. Lancet 2004; 364: 1321-8.

80. *Shoemaker WC, Appel PL, Kram HB, Waxman K, Lee TS*. Prospective trial of supranormal values of survivors as therapeutic goals in high-risk surgical patients. Chest 1988; 94: 1176-86.

81. *Rajaram SS, Desai NK, Kalra A, et al*. Pulmonary artery catheters for adult patients in intensive care. Cochrane Database Syst Rev 2013; DOI: https://doi.org/10.1002/14651858.CD003408.pub3.

82. *Sort P, Navasa M, Arroyo V, et al*. Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. N Engl J Med 1999; 341: 403-9.

83. *Finfer S, Bellomo R, Boyce N, et al*. A comparison of albumin and saline for fluid resuscitation in the intensive care unit. N Engl J Med 2004; 350: 2247-56.

84. *Davis RF, Lappas DG, Kirklin JK, Buckley MJ, Lowenstein E*. Acute oliguria after cardiopulmonary bypass: renal functional improvement with low-dose dopamine infusion. Crit Care Med 1982; 10: 852-6.

85. *Kellum JA, Decker JM*. Use of dopamine in acute renal failure: a meta-analysis. Crit Care Med 2001; 29: 1526-31.

86. *van den Berghe G, Wouters P, Weekers F, et al*. Intensive insulin therapy in critically ill patients. N Engl J Med 2001; 345: 1359-67.

87. *NICE-SUGAR Study Investigators; Finfer S, Chittock DR, Su SY, et al*. Intensive versus conventional glucose control in critically ill patients. N Engl J Med 2009; 360: 1283-97.

88. *Rivers E, Nguyen B, Havstad S, et al*. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001; 345: 1368-77.

89. *Angus DC, Barnato AE, Bell D, et al*. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. Intensive Care Med 2015; 41: 1549-60.

90. *Mangano DT, Layug EL, Wallace A, Tateo I*. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. Multicenter Study of Perioperative Ischemia Research Group. N Engl J Med 1996; 335: 1713-20.

91. *POISE Study Group; Devereaux PJ, Yang H, Yusuf S, et al*. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. Lancet 2008; 371: 1839-47.

92. *Myles PS, Leslie K, McNeil J, Forbes A, Chan MT*. Bispectral index monitoring to prevent awareness during anaesthesia: the B-Aware randomised controlled trial. Lancet 2004; 363: 1757-63.

93. *Avidan MS, Zhang L, Burnside BA, et al*. Anesthesia awareness and the bispectral index. N Engl J Med 2008; 358: 1097-108.

94. *Sessler DI, Sigl JC, Kelley SD, et al*. Hospital stay and mortality are increased in patients having a "triple low" of low blood pressure, low bispectral index, and low minimum alveolar concentration of volatile anesthesia. Anesthesiology 2012; 116: 1195-203.

95. *Sessler DI, Turan A, Stapelfeldt WH, et al*. Triple-low alerts do not reduce mortality: a real-time randomized trial. Anesthesiology 2019; 130: 72-82.

96. *Golf M, Daniels SE, Onel E*. A phase 3, randomized, placebo-controlled trial of DepoFoam(R) bupivacaine (extended-release bupivacaine local analgesic) in bunionectomy. Adv Ther 2011; 28: 776-88.

97. *Hussain N, Brull R, Sheehy B, et al*. Perineural liposomal bupivacaine is not superior to nonliposomal bupivacaine for peripheral nerve block analgesia. Anesthesiology 2021; 134: 147-64.

98. *Bashandy GM*, *Abbas DN*. Pectoral nerves I and II blocks in multimodal analgesia for breast cancer surgery: a randomized clinical trial. Reg Anesth Pain Med 2015; 40: 68-74.

99. *Cros J*, *Senges P*, *Kaprelian S*, *et al*. Pectoral I block does not improve postoperative analgesia after breast cancer surgery: a randomized, double-blind, dual-centered controlled trial. Reg Anesth Pain Med 2018; 43: 596-604.

100. *Kim DH*, *Beathe JC*, *Lin Y*, *et al*. Addition of infiltration between the popliteal artery and the capsule of the posterior knee and adductor canal block to periarticular injection enhances postoperative pain control in total knee arthroplasty: a randomized controlled trial. Anesth Analg 2019; 129: 526-35.

101. *Patterson ME*, *Vitter J*, *Bland K*, *Nossaman BD*, *Thomas LC*, *Chimento GF*. The effect of the IPACK block on pain after primary TKA: a double-blinded, prospective, randomized trial. J Arthroplasty 2020; 35: S173-7.

102. *Cummings KC 3rd*, *Napierkowski DE*, *Parra-Sanchez I*, *et al*. Effect of dexamethasone on the duration of interscalene nerve blocks with ropivacaine or bupivacaine. Br J Anaesth 2011; 107: 446-53.

103. *Chong MA*, *Berbenetz NM*, *Lin C*, *Singh S*. Perineural versus intravenous dexamethasone as an adjuvant for peripheral nerve blocks: a systematic review and meta-analysis. Reg Anesth Pain Med 2017; 42: 319-26.

104. *Wickham H*, *Cook D*, *Hofmann H*, *Buja A*. Graphical inference for Infovis. IEEE Trans Vis Comput Graph 2010; 16: 973-9.