Over the course of a 2 day expert workshop held on the topic of cancer and anaesthesia at the College of Anaesthetists of Ireland in Dublin, Ireland, the following consensus statement was developed by delegates to this unique BJA workshop (see online BJA open access supplement July 2014). This followed from their presentations of their own ongoing cutting edge research in this area and intensive interactive discussions around the existing literature and the priorities for future research. Initial brainstorming sessions resulted in a list of statements under various categories that was progressively distilled to the following concise summary after extensive inclusive discussion. We urge national and international research funding bodies to take note of these recommendations, particularly in terms of funding large-scale prospective, randomized, blinded clinical trials that can most effectively address the important clinical questions raised. We also urge the anaesthesia and cancer research communities to comment by corresponding with the BJA through its eLetter mechanism (http://bja.oxfordjournals.org/letters/).

(i) While the concept that anaesthetic or analgesic technique might affect cancer outcomes is intriguing, there is currently insufficient evidence to support any change in clinical practice.

(ii) Available data on the role of opioids in cancer are conflicting, possibly due to the use of different experimental models. Long-term opioid administration in subanalgesic doses in mice without surgery suggests that morphine promotes cancer growth. In contrast, mouse models resembling the perioperative setting, using analgesic doses of opioid, suggest either a protective effect of opioids for cancer or no effect.

(iii) Morphine does not appear to stimulate tumour initiation, and there is currently no evidence that morphine analgesia causes cancer. Whether opioid administration augments the risk of recurrence or metastasis after cancer surgery remains unclear. Currently, available research data are insufficient to indicate a change of clinical practice.

(iv) Collaboration should be sought with other specialists (medical and surgical oncologists, cancer immunologists) and existing clinical oncology registries (e.g. European Organisation for Research and Treatment of Cancer) in an attempt to study the link between anaesthetic technique(s) and cancer outcomes in surgical oncology patients.

(v) Based on recent experimental research, the expert group calls for randomized clinical trials to evaluate the effect of adjunct medications used during anaesthesia for primary cancer surgery on cancer recurrence or metastasis. Specific recommendation for further evaluation includes the effects of regional anaesthesia and analgesia, i.v. lidocaine, and non-steroidal anti-inflammatory drugs on cancer recurrence and metastasis.

Declaration of interest

D.J.B., D.M., and H.C.H. are members of the Editorial Board and H.C.H. is one of the editors of the BJA. G.S. is a member of the Associate Board of the BJA. M.W.H. is the section editor of Anesthesia and Analgesia.

Power of negative thinking

M. S. Avidan* and T. S. Wildes

Department of Anesthesiology, Washington University School of Medicine, Campus Box 8054, 660 S Euclid Ave, St Louis, MO 63110, USA
*Corresponding author. E-mail: avidannm@anest.wustl.edu

Medical research adds value when it is associated with a meaningful reduction in uncertainty, and especially when it has relevance to clinical practice and to society. Both positive (rejection of null hypotheses) and negative (inability to reject null hypotheses) findings have the potential to reduce uncertainty and to inform best practice. Several key steps are typically followed in clinical research. These include: (i) articulation of a refutable, clinically relevant hypothesis motivated by biological plausibility or preliminary research; (ii) design of a targeted experiment with appropriate statistical methods and a suitable sample size to test the null hypothesis; (iii) specification of the setting, population, intervention, and comparator; (iv) choice of relevant outcomes with appropriate statistical methods; (v) generation of results that provide evidence endorsing or contradicting the null hypothesis; and (vi) interpretation of the results in the context of the existing evidence base. In this issue of the British Journal of Anaesthesia, Landoni and colleagues report the results of a multi-centre randomized study, which tested the following null hypothesis: compared with propofol-based anaesthesia, sevoflurane-based anaesthesia is not associated
with decreased mortality or intensive care unit stay after high-risk cardiac surgery, defined as combined valvular and coronary procedures.

There is strong biological plausibility and preliminary clinical evidence to suggest that sevoflurane is associated with myocardial protection, with the most commonly advanced protective mechanism being pharmacological cardiac preconditioning. In contrast, Landoni and colleagues found that volatile-based anaesthesia with sevoflurane was not superior to anaesthetic maintenance with propofol. This is clinically important and flies in the face of compelling laboratory evidence, of emerging expert consensus about the superiority of volatile-based anaesthesia for patients with ischaemic heart disease, and even of a Bayesian sensus about the superiority of volatile-based anaesthesia for maintenance with propofol. This is clinically important and flies in the face of existing studies, such as the current study by Landoni and colleagues, which have often altered standard-of-care, could not be replicated when verification studies were conducted. Indeed, a review last year showed that over a decade, approximately half the attempts to replicate high-profile trials published in the *New England Journal of Medicine* had either contradictory or non-corroboratory findings. The likelihood that there is a plethora of false-positive reports in the peer-reviewed literature demands sober reflection, and could provide impetus for policy makers and those drafting guidelines generally to require independent replication of positive trials before implementing changes in patient care. Another suggestion is that a more stringent P-value (e.g. <0.005 or 0.001) could decrease substantially the number of false-positive reports in the scientific literature.

In view of the probable over-representation of false-positive trials in the medical literature, the publication of negative studies, such as the current study by Landoni and colleagues, assumes a particular salience in that negative trials can inject much needed balance and caution. However, this study’s conclusions must be interpreted in the context of its small size. In-hospital mortality was prioritized as an outcome measure but, possibly considering that early postoperative mortality would be <10% even in this high-risk cohort, the investigators chose a composite primary outcome. The effect of combining mortality with prolonged intensive care unit stay (>2 days) considerably boosted the estimated incidence of the outcome. A major motivation for a choosing a composite outcome is that fewer patients can be enrolled to an appropriately powered clinical trial, rendering the study economically and logistically feasible. Unfortunately, when the outcomes being grouped carry markedly different connotations, the result may be a dilution of the study’s clinical importance and interpretability.

It is conceivable that some of the published laboratory, translational, and clinical studies regarding the hypothesized superiority of volatile-based anaesthesia have yielded false-positive results. Possible reasons include the novelty of positive studies, publication bias, and plausible statistical explanations for over-representation of false-positive studies in the literature. To illustrate one statistical pitfall, consider 1000 untested hypotheses with an estimated *a priori* probability of 10% of being true, such as that several drugs are associated with a reduction in postoperative mortality. If various studies testing these hypotheses were carried out, designed with 80% power and a significance level of 0.05 as is typical, a positive study result with Drug X could surprisingly have an estimated 36% chance of being a false-positive result (i.e. a finding of a mortality benefit with Drug X even if none existed) (Fig. 1). Even a conservative analysis focusing on the highest impact medical journals suggests that >14% of positive studies could yield false-positive results. It is therefore unsurprising that several high-profile positive trials, which have often altered standard-of-care, could not be replicated when verification studies were conducted. Indeed, a review last year showed that over a decade, approximately half the attempts to replicate high-profile trials published in the *New England Journal of Medicine* had either contradictory or non-corroboratory findings. The likelihood that there is a plethora of false-positive reports in the peer-reviewed literature demands sober reflection, and could provide impetus for policy makers and those drafting guidelines generally to require independent replication of positive trials before implementing changes in patient care. Another suggestion is that a more stringent P-value (e.g. <0.005 or 0.001) could decrease substantially the number of false-positive reports in the scientific literature.

In view of the probable over-representation of false-positive trials in the medical literature, the publication of negative studies, such as the current study by Landoni and colleagues, assumes a particular salience in that negative trials can inject much needed balance and caution. However, this study’s conclusions must be interpreted in the context of its small size. In-hospital mortality was prioritized as an outcome measure but, possibly considering that early postoperative mortality would be <10% even in this high-risk cohort, the investigators chose a composite primary outcome. The effect of combining mortality with prolonged intensive care unit stay (>2 days) considerably boosted the estimated incidence of the outcome. A major motivation for a choosing a composite outcome is that fewer patients can be enrolled to an appropriately powered clinical trial, rendering the study economically and logistically feasible. Unfortunately, when the outcomes being grouped carry markedly different connotations, the result may be a dilution of the study’s clinical importance and interpretability. Additionally, in this study, the investigators assumed a high 60% baseline incidence of the composite primary outcome, and designed the trial to detect a 20% absolute reduction in this outcome solely based on the anaesthetic technique (sevoflurane-based vs propofol-based), which might be considered over-optimistic. Such a design could be characterized as underpowered or, as the authors themselves caution,
‘vulnerable to type II errors’, and the results were predictably imprecise. This study did not find a statistically significant difference between groups in the primary outcome. But the estimated 5% reduction in the primary outcome in the propofol group could well be viewed as clinically important, and the 95% confidence interval for this estimate (−9 to 19%) almost spans the investigators’ pre-specified minimum clinically important difference of 20%.

Small studies yield imprecise findings (i.e. the confidence interval around the estimated effect size is wide), and are therefore frequently not associated with a meaningful reduction in uncertainty. A strong view has been expressed that underpowered studies are unethical because they consume substantial resources, patients unknowingly participate in exercises that cannot yield definitive findings, the results might go unpublished, and they are arguably scientifically useless.14 An alternative perspective is that a single study is seldom definitive, and small trials (whether positive or negative) can generate important hypotheses, provide estimates of effect sizes, and can be included in meta-analyses.15 Our opinion is that it is obviously preferable to conduct studies that are designed to yield precise and meaningful results, but once a study is completed, as long as it is methodologically sound, it should be published in the scientific literature, and its value can be assessed in the context of other relevant publications.

The third explanation for the unexpected results of the Landoni study is that volatile-based anaesthesia might be superior for some patients with ischaemic heart disease undergoing certain surgical procedures but not for other surgical patients. This account is challenging from a philosophical standpoint. Popper16 argued that a distinguishing feature of a scientific theory is that an experiment can be designed to attempt falsification. For example, if we theorize that the administration of propofol to animals results in loss of responsiveness, we can readily design experiments to attempt falsification of this premise. On the other hand, if we hypothesize that volatile anaesthetics are associated with improved clinical outcomes compared with i.v. anaesthetics in a subgroup of patients with certain known characteristics (e.g. ischaemic heart disease) who are undergoing specific surgical procedures (e.g. on-pump coronary artery bypass surgery), and who lack characteristics that negate the beneficial effect (e.g. specific comorbidities or genetic risk factors), it is considerably harder to design an experiment that could reliably falsify such a convoluted proposition.

In assessing a study, an important question is whether its design truly has the potential to falsify a theory.

In the final analysis, the paper by Landoni and colleagues is valuable in that it plausibly challenges the established paradigm, which is all the more notable in that the same group has previously contributed research that has reinforced the proposition that volatile-based anaesthesia is cardio-protective. As stated by Popper, ‘For if we are uncritical we shall always find what we want: we shall look for, and find, confirmations, and we shall look away from, and not see, whatever might be dangerous to our pet theories. In this way it is only too easy to obtain what appears to be overwhelming evidence in favour of a theory which, if approached critically, would have been refuted. In order to make the method of selection by elimination work, and in order to ensure that only the fittest theories survive, their struggle for life must be made severe for them’.16

Declaration of interest

M.S.A. is a member of the Associate Editorial Board of the BJA.

References

2 Parfrey P, Ravani P. On framing the research question and choosing the appropriate research design. Methods Mol Biol 2009; 473: 1 – 17