EDITORIAL

Randomized clinical trials and neurosurgery

Fred G. Barker II, MD

Department of Neurosurgery, Massachusetts General Hospital, Boston, Massachusetts

IN 1996, *Lancet* editor Richard Horton angered a generation of academic surgeons by equating surgical research with “comic opera.”²⁰ He challenged the field to improve; and an important part of this, he believed, would be for surgeons to turn away from the case series that had traditionally clogged our journals toward performing more randomized controlled trials (RCTs). In this issue of the *Journal*, Mansouri et al. report on their examination of the quality of reporting of a cohort of RCTs on neurosurgical procedures published between 2000 and 2014.²⁹ Sixty-one RCTs published in a variety of neurosurgical and nonneurosurgical journals were found. The most frequent subject domains were vascular, functional, and tumor neurosurgery; spine surgery RCTs were excluded by design. Overall, the reporting quality of the RCTs was poor, particularly for those studies that were published in neurosurgical journals. Major deficiencies were present in reporting of blinding, allocation concealment, justification of sample size, and to a lesser degree, in completeness of subject follow-up. The number of RCTs published per year was low in comparison with other medical specialties. RCTs had high impact: on average they were cited more than 10 times per year since publication. The authors’ major conclusion was that neurosurgeons should explore alternate means of answering clinical questions, specifically registry studies.

The authors’ conclusions about the numbers of RCTs on neurosurgical topics and their quality of reporting are likely to be sound, both because of the careful methods used and because their findings are consistent with similar previously published studies. In 2004, Vranos et al.⁵² found 108 RCTs published during 1966–2004 that contained a neurosurgical procedure in at least 1 trial arm; approximately half addressed spinal surgery, so the rate of publication was less than 3 RCTs per year (slightly more than 1 per year, excluding spine studies). Mansouri et al. found an average of 5 RCTs per year from 2000 to 2014. In a study limited to 3 major neurosurgical journals, Gnanalingham et al.¹⁶ found a modest increase in RCTs

reported, from 0% of all articles in 1982 to 2% in 2002, a trend confirmed by a 1966–2006 survey that used EMBASE as well as MEDLINE and was not limited to neurosurgical journals.⁴³ Although Mansouri et al. found no increasing trend in numbers of neurosurgical RCTs between 2000 and 2014, the longer-term trend does appear to be positive. This would be good news, because presently neurosurgery seems to have fewer RCTs in its literature and knowledge base than almost any other medical or surgical specialty.⁵⁶ For example, a study on major orthopedic journals found 4%–6% of all articles to be RCTs (2006–2010),⁶ and total numbers of orthopedic RCTs found in a systematic review (2008–2011) were 40 per year.⁴⁷ Cardiothoracic surgery journals contain about 6% RCTs, general surgery journals about 7%, and general medical journals about 12%.^{20,49} In specialty spine journals the rate is more than 4%.¹⁵ One consequence of this neurosurgical poverty is that our “meta-analyses” are less likely to include data from RCTs than those in any other specialty.⁴¹ The same is probably true for our daily clinical decisions.

So neurosurgical RCTs are rare birds; but can we perhaps be happy with their quality, if not their quantity? Here, too, we will be disappointed. Neurosurgical RCTs had low quality scores on two widely-used metrics, especially when published in our own journals.²⁹ Again, while direct cross-study comparisons are not possible for these scores, due to subjectivity in the scoring scales and incomplete reporting of results, prior studies of RCTs in neurosurgery^{23,28} and spine surgery³⁵ also showed major deficiencies in RCT quality measures. In some contexts, it has been shown that “bad reporting does not mean bad methodology,” i.e., authors used high quality methods in their trials but failed to report them.^{32,46} Notably, though, these studies originated in US Cooperative Cancer Groups in which tight supervision of trial quality is structurally guaranteed. It is more likely that the deficiencies noted by Mansouri et al. are real, but journals could address this by publishing trial protocols as a supplement to trial reports. Still worse news is that Mansouri et al. scored neurosurgical

RCTs using a scale based on the standard Consolidated Standards for Reporting of Trials (CONSORT) criteria,³³ not the more rigorous nonpharmacological extension specifically designed for trials that evaluate procedures.⁴ If the nonpharmacological scale had been used, the deficiencies found would likely have been even deeper and more widespread.^{34,45} The extra criteria in the CONSORT extension address factors such as skill of participating surgeons and completeness of description of the procedure studied, and any measures of compliance with the protocol definition of the procedure.^{22,50} The deficiencies found by Mansouri and coauthors in neurosurgical RCTs affected blinding (adequately performed and reported in less than 10% of neurosurgical RCTs), adequate concealment of treatment allocation (important to prevent investigators from cheating on randomization; adequate in 23%–50% of trials), and sample size calculation (adequate in 20%–38% of trials). There is objective evidence that failures in blinding and allocation concealment are associated with biased results of RCTs.⁴²

Factors that make surgical RCTs more difficult to perform than medical trials are well known, and some of these factors can help explain this study's findings. Blinding is a particular challenge in surgical trials and had the lowest compliance rate of the CONSORT criteria in this RCT cohort. Yet surgery has a powerful placebo effect that should not be ignored in trial design, particularly for procedures that are less invasive than craniotomy.⁵³ In Parkinson's disease, surgery had a placebo effect almost 5 times stronger than medications, and similar findings have been reported in migraine.^{17,30} Various techniques have been used to address this key difficulty. The most broadly applicable technique is to blind the outcome assessor to surgical status, as successfully used in the North American Symptomatic Carotid Endarterectomy Trial (NASCET).³⁷ For trials testing electronic implants, staggered activation and/or alternating between on- and off-status in a blinded fashion is commonly used.²⁷ And for relatively noninvasive surgery, actual sham procedures, with all of their ethical challenges, have been recommended.^{12,40} In surgical trials, standardizing the procedure and those who perform it is of special importance, due to factors such as the surgical learning curve and the volume-outcome effect.⁴ The same considerations make the choice of when to initiate a surgical trial complicated: too early, and the procedure may not be fully developed, too late, and it may already be considered "standard practice" and impossible to deny to patients in a trial. For allocation concealment and sample size calculation reporting, areas of particularly poor performance in this RCT cohort, it is hard to escape the conclusion that the deficiencies are the result of inexperienced neurosurgical trialists, authors, reviewers, editors, and—in the final analysis—readers.

What to do? I would emphasize different solutions from those highlighted by Mansouri et al. There is no doubt that we can, and will, learn a great deal about many routine neurosurgical procedures through important prospective registry efforts now being started, such as the National Neurosurgery Quality and Outcomes Database (N²QOD) registry, which is sponsored by multiple neurosurgical professional societies.¹ For many important goals—studying

the natural history of a condition, defining prognostic factors, finding variation in the use or quality of delivery of an established treatment, monitoring safety, harm, and the success of generalizing a treatment to the broader community—registries are clearly superior to RCTs, except in so far as every RCT is also a registry.¹⁴ However, registries capture results of standard medical practice, in which patient and physician choices determine treatment selection. This complex process invariably creates treatment and comparison "control" groups that are imbalanced for every variable we could possibly measure (as statistician Marks Nester's "applied statistician's creed" reminds us, "no two populations are identical in any respect."³⁶) Some of these baseline imbalances can be ignored, because they are in variables that are not prognostic for outcome (for example, zodiac sign).^{2,21} Some are known prognostic variables, and we can plan ahead in a registry to collect the necessary information to adjust our outcomes using multivariate risk adjustment or propensity scores^{3,31,54} to make a fair treatment comparison.

Treatment groups in registries are also unbalanced for unknown prognostic factors as well, and statistical techniques cannot adjust these differences away (unless the unknown confounders are perfectly correlated with known factors, an assumption too strong to be plausible). For example, patients with glioblastoma who are married live about 20% longer, and those undergoing both surgery and radiation are more likely to be married;⁵ living more than 15 miles from the treatment center carried a risk of death about two-thirds lower than those who lived closer in a review of cancer treatment trials, largely using ineffective drugs;²⁵ and eligibility for treatment accounted entirely for the apparent treatment benefit of brachytherapy for glioblastoma when finally tested using RCTs.^{11,26,44} Such prognostic factors can be too difficult to define or too expensive to archive in a registry, or so unexpected that we do not consider collecting them at all. The difficulty of defining some hidden confounders makes them practically intangible. For example, patients who adhere better to prescribed placebos—ineffective by definition—have significantly lower rates of cardiac mortality in myocardial infarction prevention trials, less bone loss, lower rates of hip fracture in fracture prevention trials, and lower mortality from cancer and from all causes in hormone replacement trials.^{8–10,19} How can we measure this quality in a surgical registry? Randomization, however, balances every prognostic factor, both known and unknown. This is why effective randomization is of paramount importance in an unbiased treatment comparison.

Some additional problems with neurosurgical RCTs, including some of those identified by Mansouri et al., are not really avoided in registry studies either. For example, if blinding and allocation concealment are poorly conducted in neurosurgical RCTs, they are never part of a registry study at all. Crossovers in an RCT have their equivalent in every patient in a registry, since treatment selection by patient and physician choice is the norm in standard practice. Even the difficulty establishing equipoise necessary for an RCT is not such a clear advantage for registries. If there truly is no equipoise in a community for a treatment decision, registry comparisons will be "apples

to oranges” and hence invalid. When there is equipoise around a treatment decision in a community that would allow a valid registry comparison, it is our own fault if we cannot find surgical investigators who acknowledge that colleagues are making a different treatment choice from their own, and that randomization might be ethical—and find ways of convincing patients of this as well.¹³

It is often said that neurosurgical RCTs are the ultimate in enrollment challenges, because neurosurgery (spine excepted) is a collection of rare diseases in which a mistake in treatment can be crippling or even fatal. This is even more true of pediatric oncology, and yet 71% of children with cancer in the US are enrolled in clinical trials (compared with about 5% for adults).⁴⁸ On the border between childhood and adulthood, US adolescent patients with cancer can be evaluated by either a pediatric oncologist or an adult oncologist. Seventy-one percent of those seen by pediatric oncologists are enrolled in clinical trials, compared with 11% of those seen by adult specialists, an odds ratio in multivariate analysis of 7.4 ($p < 0.001$).³⁸ For US adult patients with cancer, medical oncologists are 5 times more likely than radiation oncologists to enroll patients in available trials; surgical oncologists are 4 times more likely to enroll patients than general surgeons.²⁴ Medical culture matters in determining whether patients enter clinical trials.

My response to this well-conducted and timely study, then, would be to suggest we focus our attention on understanding why neurosurgeons do so poorly with RCTs, and how we can change the game. In 2015, no form of medical publication rivals the RCT for academic and “real world” clinical impact. Mansouri et al. found a very high citation index for the RCTs they studied, about 3 times that for an average paper in our journals. Surveys show that neurosurgeons strongly prefer RCTs as a source of data on treatment comparisons: in knowledge gap surveys conducted before the 2007 Congress of Neurological Surgeons Annual Meeting, 77% of neurosurgeons preferred RCT evidence for a question about brain metastasis treatment, compared with just 4% who wanted to see data from outcome registries; for aneurysm clipping versus coiling, preferences were 67% for RCTs and 5% for registries.^{7,51} Properly conducted and reported RCTs change practice rapidly and profoundly, as has been shown time and again in population-based before-and-after studies.^{18,39,55} We have not yet seen this from registry research. Mansouri and colleagues are probably right in their conclusions that neurosurgeons produce few RCTs, and those of low quality; we should thank them for the demonstration. But rather than turn away from such a powerful tool, we should instead set about learning how to use it. In Harvey Cushing’s day, learning the subtleties of neurological diagnosis was the challenge for neurosurgeons, so they could choose their own operations and work to improve them. For our generation, learning how to master clinical studies—of every kind—is the equivalent challenge.

<http://thejns.org/doi/abs/10.3171/2015.2.JNS142960>

References

1. Asher AL, Speroff T, Dittus RS, Parker SL, Davies JM, Selden N, et al: The National Neurosurgery Quality and

- Outcomes Database (N2QOD): a collaborative North American outcomes registry to advance value-based spine care. *Spine (Phila Pa 1976)* **39** (22 Suppl 1):S106–S116, 2014
2. Austin PC, Mamdani MM, Juurlink DN, Hux JE: Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *J Clin Epidemiol* **59**:964–969, 2006
3. Barker FG II, Chang SM, Gutin PH, Malec MK, McDermott MW, Prados MD, et al: Survival and functional status after resection of recurrent glioblastoma multiforme. *Neurosurgery* **42**:709–723, 1998
4. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P: Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med* **148**:295–309, 2008
5. Chang SM, Barker FG II: Marital status, treatment, and survival in patients with glioblastoma multiforme: a population based study. *Cancer* **104**:1975–1984, 2005
6. Chess LE, Gagnier J: Risk of bias of randomized controlled trials published in orthopaedic journals. *BMC Med Res Methodol* **13**:76, 2013
7. Connolly ES Jr, Hoh BL, Selden NR, Asher AL, Kondziolka D, Boulis NM, et al: Clipping versus coiling for ruptured intracranial aneurysms: integrated medical learning at CNS 2007. *Neurosurgery* **66**:19–34, 2010
8. Coronary Drug Project Research Group: Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* **303**:1038–1041, 1980
9. Curtis JR, Delzell E, Chen L, Black D, Ensrud K, Judd S, et al: The relationship between bisphosphonate adherence and fracture: is it the behavior or the medication? Results from the placebo arm of the fracture intervention trial. *J Bone Miner Res* **26**:683–688, 2011
10. Curtis JR, Larson JC, Delzell E, Brookhart MA, Cadarette SM, Chlebowski R, et al: Placebo adherence, clinical outcomes, and mortality in the women’s health initiative randomized hormone therapy trials. *Med Care* **49**:427–435, 2011
11. Florell RC, Macdonald DR, Irish WD, Bernstein M, Leibel SA, Gutin PH, et al: Selection bias, survival, and brachytherapy for glioma. *J Neurosurg* **76**:179–183, 1992
12. Freeman TB, Vawter DE, Leaverton PE, Godbold JH, Hauser RA, Goetz CG, et al: Use of placebo surgery in controlled trials of a cellular-based therapy for Parkinson’s disease. *N Engl J Med* **341**:988–992, 1999
13. Ghogawala Z, Benzel EC, Amin-Hanjani S, Coumans JV, Sonntag VK, Borges LF, et al: Spinal experts’ review significantly increases patient consent to randomization in the SLIP study: an ongoing prospective, multicenter surgical trial. *Neurosurgery* **57**:398, 2005 (Abstract)
14. Glicklich RE, Dreyer NA, Leavy MB (eds): **Registries for Evaluating Patient Outcomes: A User’s Guide**. Rockville, MD: Agency for Healthcare Research and Quality, 2014, Vol 1
15. Gnanalingham KK, Davies BM, Balamurali G, Titoria P, Doyle P, Abou-Zeid A: Improving levels of evidence in studies published in spinal journals from 1983 to 2011. *Br J Neurosurg* **27**:152–155, 2013
16. Gnanalingham KK, Tysome J, Martinez-Canca J, Barazi SA: Quality of clinical studies in neurosurgical journals: signs of improvement over three decades. *J Neurosurg* **103**:439–443, 2005
17. Goetz CG, Wu J, McDermott MP, Adler CH, Fahn S, Freed CR, et al: Placebo response in Parkinson’s disease: comparisons among 11 trials covering medical and surgical interventions. *Mov Disord* **23**:690–699, 2008
18. Gross CP, Steiner CA, Bass EB, Powe NR: Relation between prepublication release of clinical trial results and the practice of carotid endarterectomy. *JAMA* **284**:2886–2893, 2000

19. Hartz A, He T: Why is greater medication adherence associated with better outcomes. **Emerg Themes Epidemiol** 10:1, 2013
20. Horton R: Surgical research or comic opera: questions, but few answers. **Lancet** 347:984–985, 1996
21. ISIS-2 Collaborative Group: Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. **Lancet** 2:349–360, 1988
22. Jacquier I, Boutron I, Moher D, Roy C, Ravaud P: The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. **Ann Surg** 244:677–683, 2006
23. Kiehna EN, Starke RM, Pouratian N, Dumont AS: Standards for reporting randomized controlled trials in neurosurgery. **J Neurosurg** 114:280–285, 2011
24. Klabunde CN, Keating NL, Potosky AL, Ambis A, He Y, Hornbrook MC, et al: A population-based assessment of specialty physician involvement in cancer clinical trials. **J Natl Cancer Inst** 103:384–397, 2011
25. Lamont EB, Hayreh D, Pickett KE, Dignam JJ, List MA, Stenson KM, et al: Is patient travel distance associated with survival on phase II clinical trials in oncology? **J Natl Cancer Inst** 95:1370–1375, 2003
26. Laperriere NJ, Leung PM, McKenzie S, Milosevic M, Wong S, Glen J, et al: Randomized study of brachytherapy in the initial management of patients with malignant astrocytoma. **Int J Radiat Oncol Biol Phys** 41:1005–1011, 1998
27. Little S, Pogossyan A, Neal S, Zavala B, Zrinzo L, Hariz M, et al: Adaptive deep brain stimulation in advanced Parkinson disease. **Ann Neurol** 74:449–457, 2013
28. Lu J, Gary KW, Copolillo A, Ward J, Niemeier JP, Lapane KL: Randomized controlled trials in adult traumatic brain injury: a review of compliance to CONSORT Statement. **Arch Phys Med Rehabil** 96:702–714, 2015
29. Mansouri A, Cooper B, Shin SM, Kondziolka D: Randomized controlled trials and neurosurgery: the ideal fit or should alternative methodologies be considered? **J Neurosurg** [epub ahead of print August 28, 2015. DOI: 10.3171/2014.12.JNS142465]
30. Meissner K, Fässler M, Rücker G, Kleijnen J, Hróbjartsson A, Schneider A, et al: Differential effectiveness of placebo treatments: a systematic review of migraine prophylaxis. **JAMA Intern Med** 173:1941–1951, 2013
31. Mertens BJ, Jacobs WC, Brand R, Peul WC: Assessment of patient-specific surgery effect based on weighted estimation and propensity scoring in the re-analysis of the sciatica trial. **PLoS ONE** 9:e111325, 2014
32. Mhaskar R, Djulbegovic B, Magazín A, Soares HP, Kumar A: Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. **J Clin Epidemiol** 65:602–609, 2012
33. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. **BMJ** 340:c869, 2010
34. Nagendran M, Harding D, Teo W, Camm C, Maruthappu M, McCulloch P, et al: Poor adherence of randomised trials in surgery to CONSORT guidelines for non-pharmacological treatments (NPT): a cross-sectional study. **BMJ Open** 3:e003898, 2013
35. Naunheim MR, Walcott BP, Nahed BV, Simpson AK, Agarwalla PK, Coumans JV: The quality of randomized controlled trial reporting in spine literature. **Spine (Phila Pa 1976)** 36:1326–1330, 2011
36. Nester MR: An applied statistician's creed. **Appl Stat** 45:401–410, 1996
37. North American Symptomatic Carotid Endarterectomy Trial Collaborators: Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. **N Engl J Med** 325:445–453, 1991
38. Parsons HM, Harlan LC, Seibel NL, Stevens JL, Keegan TH: Clinical trial participation and time to treatment among adolescents and young adults with cancer: does age at diagnosis or insurance make a difference? **J Clin Oncol** 29:4045–4053, 2011
39. Qureshi AI, Vazquez G, Tariq N, Suri MF, Lakshminarayan K, Lanzino G: Impact of International Subarachnoid Aneurysm Trial results on treatment of ruptured intracranial aneurysms in the United States. Clinical article. **J Neurosurg** 114:834–841, 2011
40. Redberg RF: Sham controls in medical device trials. **N Engl J Med** 371:892–893, 2014
41. Sampson JH, Barker FG II: Methodology and reporting of meta-analyses in the neurosurgical literature. **J Neurosurg** 120:791–794, 2014
42. Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al: Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. **Ann Intern Med** 157:429–438, 2012
43. Schöller K, Licht S, Tonn JC, Uhl E: Randomized controlled trials in neurosurgery—how good are we? **Acta Neurochir (Wien)** 151:519–527, 2009
44. Selker RG, Shapiro WR, Burger P, Blackwood MS, Arena VC, Gilder JC, et al: The Brain Tumor Cooperative Group NIH Trial 87-01: a randomized comparison of surgery, external radiotherapy, and carmustine versus surgery, interstitial radiotherapy boost, external radiation therapy, and carmustine. **Neurosurgery** 51:343–357, 2002
45. Sinha S, Ashby E, Jayaram R, Grocott MP: Quality of reporting in randomized trials published in high-quality surgical journals. **J Am Coll Surg** 209:565–571, 2009
46. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al: Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. **BMJ** 328:22–24, 2004
47. Somerson JS, Bhandari M, Vaughan CT, Smith CS, Zelle BA: Lack of diversity in orthopaedic trials conducted in the United States. **J Bone Joint Surg Am** 96:e56, 2014
48. Tejeda HA, Green SB, Trimble EL, Ford L, High JL, Ungerleider RS, et al: Representation of African-Americans, Hispanics, and whites in National Cancer Institute cancer treatment trials. **J Natl Cancer Inst** 88:812–816, 1996
49. Tiruvoipati R, Balasubramanian SP, Atturu G, Peek GJ, Elbourne D: Improving the quality of reporting randomized controlled trials in cardiothoracic surgery: the way forward. **J Thorac Cardiovasc Surg** 132:233–240, 2006
50. van Oldenrijk J, van Berkel Y, Kerkhoffs GM, Bhandari M, Poolman RW: Do authors report surgical expertise in open spine surgery related randomized controlled trials? A systematic review on quality of reporting. **Spine (Phila Pa 1976)** 38:857–864, 2013
51. Vogelbaum MA, Asher AL, Kondziolka D, Boulis NM, Selden NR, Hoh BL, et al: Modern treatment of cerebral metastases: Integrated Medical Learning(SM) at CNS 2007. **J Neurooncol** 93:89–105, 2009
52. Vranos G, Tatsioni A, Polyzoidis K, Ioannidis JP: Randomized trials of neurosurgical interventions: a systematic appraisal. **Neurosurgery** 55:18–26, 2004
53. Wartolowska K, Judge A, Hopewell S, Collins GS, Dean BJ, Rombach I, et al: Use of placebo controls in the evaluation of surgery: systematic review. **BMJ** 348:g3253, 2014
54. Williamson E, Morley R, Lucas A, Carpenter J: Propensity scores: from naive enthusiasm to intuitive understanding. **Stat Methods Med Res** 21:273–293, 2012

55. Yabroff KR, Harlan L, Zeruto C, Abrams J, Mann B: Patterns of care and survival for patients with glioblastoma multiforme diagnosed during 2006. *Neuro Oncol* **14**:351–359, 2012
56. Yarascavitch BA, Chuback JE, Almenawer SA, Reddy K, Bhandari M: Levels of evidence in the neurosurgical literature: more tribulations than trials. *Neurosurgery* **71**:1131–1138, 2012

Disclosure

Dr. Barker receives salary support from the NIH Cancer Treatment Evaluation Program.

Response

Alireza Mansouri, MD,^{1,2} and Douglas Kondziolka, MD, MSc^{3,4}

¹Division of Neurosurgery, University of Toronto, Ontario, Canada;

²Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; and ³Departments of Neurosurgery and ⁴Radiation Oncology, NYU Langone Medical Center, New York University, New York, New York

Dr. Barker has elegantly highlighted the key findings of our study and their implications. While agreeing with the basis of our conclusions, he argues that registries are limited by a number of factors, which may affect their validity. Therefore, a greater emphasis should be made on identifying and addressing the root causes of deficiencies in neurosurgical RCTs rather than attempting to implement alternative parallel strategies.

Indeed, results derived from RCTs represent one of the highest qualities of evidence in medicine, second only to well-conducted meta-analyses of several RCTs. They have the potential to significantly modify practice. However, the quality of this evidence is limited by the quality of conduct and reporting of the individual RCTs themselves, let alone the fact that the method is underutilized. As summarized in our study, the pace of neurosurgical research driven by RCTs is slow and the issue of poor quality has persisted through time. Certainly, this is not limited to neurosurgery and spans the spectrum of various other surgical specialties as well. This suggests a systemic root cause. Our goal was to characterize this and attempt to propose solutions.

As noted by Dr. Barker, one of the main principles that differentiate RCTs from designs based on prospective registries is the ability to balance potential confounding factors, including those that are unknown. Essential to maintaining this balance is the fidelity of blinding which, if performed adequately, limits performance and detection bias.^{11,12} However, the observation that many neurosurgical RCTs were unable to maintain adequate blinding is a cause for concern. The lack of blinding can potentially nullify any prognostic balance attained via randomization given its impact on the practitioner's choice with regards to providing additional patient monitoring and intervention, withdrawing a patient from further intervention, and crossing a patient over from one intervention arm to another. The latter is particularly relevant for trials analyzed on an intent-to-treat basis wherein patients could “cross-

over” to the alternative treatment arm yet be analyzed as part of the original intervention arm. This was a limiting factor in analyzing the results obtained in the Spine Patient Outcomes Research Trial (SPORT) in which an intent-to-treat analysis showed no benefit between operative versus nonoperative management of lumbar disc herniation, secondary to the significant crossover between the two arms.¹⁴ Therefore, in the absence of blinding, the advantages of RCTs over registry-based trials would likely be fewer.

A further fundamental component to conducting an RCT is the issue of equipoise. Despite the presence of true equipoise within the medical community, it is not always possible to recruit individual practitioners to participate in randomized trials. The concept of surgery for adult low-grade gliomas is a prominent example whereby Sampson argues that an RCT is unlikely to ever be performed.¹⁰ In addition to personal preference/opinion, the background training of the individual practitioner can also be a limiting factor with regard to the intervention that can be offered. Through incorporating data from a variety of practices with diverse preferences, registries offer a pragmatic alternative to this obstacle. Beyond individual practitioners, the recruitment of participating patients has been a source of difficulty for RCTs as well. Numerous trials were terminated prematurely due to poor accrual,^{5,6,9,13} potentially affecting the validity of their findings or preventing meaningful analyses.² In the era of “big data” and access to the necessary statistical and computational tools to analyze them, trials based on small sample sizes could be considered an “underachievement.”

Certainly our own medical culture is a significant factor in the ability to enrol patients in trials. However, the acuity of many surgical cases is an important factor as well. In these scenarios it is particularly challenging, although certainly not impossible, to convince patients (or their practitioners) to participate in trials. Furthermore, the complexities associated with standardization of surgical techniques must also be taken into consideration. Expertise-based trials, guided by an established procedural protocol, have been posed as a potential solution; the NASCET trial is an excellent example.⁷ However, it has been argued that such rigid protocols do not reflect true practice, hence potentially limiting the generalizability of such results.

Despite the aforementioned limitations affecting RCTs, high-impact studies have indeed influenced neurosurgical practice. However, we would like to caution the reader to be mindful of potential analytical issues with such trials that can potentially misguide practice if not interpreted correctly. The brain metastasis trial by Patchell et al. that compared resection combined with whole brain radiotherapy to whole brain radiotherapy alone found a benefit to resection and changed practice broadly.⁸ However there was only 1 patient with melanoma and 1 with breast cancer in 1 of the arms. In 1990, this may have seemed unimportant because the disease being studied was “single brain metastasis” and not melanoma or breast cancer. In 2015, this deficiency is glaring. In a different example, the Carotid Revascularization Endarterectomy versus Stenting Trial (CREST), using a composite outcome of death, stroke, and myocardial infarction, showed a lack of benefit

of carotid artery stenting versus endarterectomy at 4 years of follow-up;¹ the use of this composite outcome allowed for a reduction of the required sample size. However, in selecting this composite, the authors had not considered that patients value the impact of stroke on quality of life differently from the other outcomes. Furthermore, the direction of effect on stroke and myocardial infarct was different between the two interventions and this was nullified through the use of a composite outcome incorporating both variables.

The multicenter International Subarachnoid Aneurysm Trial (ISAT), one of the largest such studies to provide a comparison of coiling versus clipping of ruptured aneurysms, was randomized by design.⁴ However, 70% of potentially eligible patients were not enrolled in the study and 95% of patients were recruited from European centers. Furthermore, while participating endovascular surgeons were required to demonstrate expertise, this was not required from surgeons performing craniotomies for clip placement. Therefore, while the trial concluded that the 1-year rate of death/dependence was lower in the coiling group, the trial's limitations must be considered prior to outlining the risks and benefits to patients. Lastly, the issue of conclusions derived from subgroup analyses must be noted. Based on a subgroup analysis, the STICH trial demonstrated a favorable outcome following surgery for hematomas less than 1 cm from the cortical surface.³ While this would be a conclusion that could be clinically reasonable, the reader must be mindful that subgroup analyses are not powered to allow for definitive conclusions. These analytic issues are certainly not restricted to RCTs. The concern, however, relates to potentially deriving erroneous conclusions and dismissing those from well-conducted observational trials simply because RCTs are trusted to present the highest quality of evidence.

Many of the issues pertaining to RCTs alluded to here are well-known. It is not our intention to detract from the value of RCTs. However, recognizing the position of RCTs within the hierarchy of evidence-based medicine and that they are the preferred source of evidence for individuals in our specialty, our aim was to highlight their deficiencies and alert readers to the potential erroneous conclusions that could be derived from poorly conducted studies. Certainly an enhanced understanding of trial conduct by neurosurgeons, close collaboration with biostatisticians/clinical epidemiologists, and more stringent reporting criteria by journals would allow for improvement. In this study, however, rather than recapitulate suggestions formulated in prior reviews of a similar nature, we sought to recognize that limitations to conducting RCTs for particular clinical queries can and will continue to exist. Neurosurgeons should not wait for the occasional RCT to be published. We should participate in broad data collection using real-world clinical practice.

Thus, our ultimate goal was to propose an alternative strategy for these specific scenarios. There is no doubt that registries are affected by various challenges and cannot match the quality of properly conducted RCTs. However, dismissing the potential benefits that can be derived from registries is likely a key factor in why they have failed to produce results that can change practice. Embracing the

reality and devising standardized protocols for the conduct of high quality, prospective, registry-based studies in parallel with RCTs has the potential to yield valid results and accelerate the advancement of knowledge in our field. We need to foster and enhance clinical research along multiple fronts. A new culture of data collection in clinical practice using robust new tools will be an important component.

References

1. Brott TG, Hobson RW II, Howard G, Roubin GS, Clark WM, Brooks W, et al: Stenting versus endarterectomy for treatment of carotid-artery stenosis. **N Engl J Med** **363**:11–23, 2010
2. Engel J Jr: The timing of surgical intervention for mesial temporal lobe epilepsy: a plan for a randomized clinical trial. **Arch Neurol** **56**:1338–1341, 1999
3. Mendelow AD, Gregson BA, Fernandes HM, Murray GD, Teasdale GM, Hope DT, et al: Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial. **Lancet** **365**:387–397, 2005
4. Molyneux A, Kerr R, Stratton I, Sandercock P, Clarke M, Shrimpton J, et al: International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomized trial. **J Stroke Cerebrovasc Dis** **11**:304–314, 2002
5. Muacevic A, Wowra B, Siefert A, Tonn JC, Steiger HJ, Kreth FW: Microsurgery plus whole brain irradiation versus Gamma Knife surgery alone for treatment of single metastases to the brain: a randomized controlled multicentre phase III trial. **J Neurooncol** **87**:299–307, 2008
6. Neuhaus T, Ko Y, Muller RP, Grabenbauer GG, Hedde JP, Schueller H, et al: A phase III trial of topotecan and whole brain radiation therapy for patients with CNS-metastases due to lung cancer. **Br J Cancer** **100**:291–297, 2009
7. North American Symptomatic Carotid Endarterectomy Trial Collaborators: Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. **N Engl J Med** **325**:445–453, 1991
8. Patchell RA, Tibbs PA, Walsh JW, Dempsey RJ, Maruyama Y, Kryscio RJ, et al: A randomized trial of surgery in the treatment of single metastases to the brain. **N Engl J Med** **322**:494–500, 1990
9. Roos DE, Wirth A, Burmeister BH, Spry NA, Drummond KJ, Beresford JA, et al: Whole brain irradiation following surgery or radiosurgery for solitary brain metastases: mature results of a prematurely closed randomized Trans-Tasman Radiation Oncology Group trial (TROG 98.05). **Radiother Oncol** **80**:318–322, 2006
10. Sampson JH: Low-grade glioma. **J Neurosurg** **114**:563–565, 2011
11. Schulz KF, Grimes DA: Blinding in randomised trials: hiding who got what. **Lancet** **359**:696–700, 2002
12. Schulz KF, Grimes DA, Altman DG, Hayes RJ: Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. **BMJ** **312**:742–744, 1996
13. Verger E, Gil M, Yaya R, Viñolas N, Villa S, Pujol T, et al: Temozolomide and concomitant whole brain radiotherapy in patients with brain metastases: a phase II randomized trial. **Int J Radiat Oncol Biol Phys** **61**:185–191, 2005
14. Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, et al: Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. **JAMA** **296**:2441–2450, 2006