



## EDITORIAL

**Patient validation of retrospective data****Ranjith Babu, MD, MS, and John H. Sampson, MD, PhD, MBA**

Duke Brain Tumor Immunotherapy Program and Preston Robert Tisch Brain Tumor Center, Division of Neurosurgery, Department of Surgery, Duke University Medical Center, Durham, North Carolina

**I**N this issue of the *Journal of Neurosurgery*, Drewes and colleagues from Norway compare outcomes assessments from retrospective reviews of hospital records to structured phone interviews conducted 30 days postoperatively in almost 200 patients who underwent resection of intracranial tumors.<sup>4</sup> This is an important paper because outcomes reports in the neurosurgical literature frequently rely on retrospective review of hospital records. Unfortunately, how accurate these are and how representative they are of the patient experience is not clear. While these authors found that the specificity of record review was high, suggesting that new deficits were unlikely to be present in the hospital records when they were not acknowledged by the patient, the sensitivity was disappointingly low, ranging from 0.07 to 0.52 across different neurological domains. This lack of sensitivity to neurological deficits resulting from neurosurgical treatment should not be unexpected. It has been well shown that adverse events are typically higher in randomized controlled trials (RCTs), where specific efforts are made to document outcomes, than in retrospective case series.<sup>3,6</sup> This study demonstrates the unreliability of such retrospective reviews and calls for validated measures of neurosurgical morbidity.

The primary use of retrospective studies as evidence for procedural efficacy and patient outcomes has been a major limitation of the neurosurgical literature. These studies, which provide only Level III, IV, or V evidence account for 89.8% of all studies in top neurosurgery journals.<sup>8</sup> Such studies, while easy to perform, suffer from numerous limitations, including selection bias, missing data, and reliance on the accuracy of clinical records. As a result, these studies have been shown to be significantly flawed with respect to the reporting of clinical diagnoses, procedure types and frequency, and complication rates.<sup>2,7</sup> With increasing awareness of the limitations of retrospective reviews, steps towards improving the quality of neurosurgery research have been made through the establishment

of prospective registries such as the National Neurosurgery Quality and Outcomes Database (N<sup>2</sup>QOD). Although data from such registries are often reported as retrospective reviews of prospectively collected data, they offer an improvement over traditional retrospective studies. Still, these studies should rarely be used to compare outcomes from different treatments unless they can demonstrate that the patient cohorts being compared were equally likely to be assigned to the different treatment arms—this is absolutely key.

An important point noted by Drewes et al. is the utility of patient-reported outcomes (PROs). These measures are reports of a patient's health status that come from the patient without further interpretation by a clinician. PROs have proven valuable in a variety of aspects of clinical research, such as adverse event monitoring and evaluation of patient quality of life. As a result, nearly a quarter of US drug products approved from 2006 to 2010 were granted treatment benefit claims based on PROs.<sup>5</sup> Due to the importance of such data, various initiatives have also been started by the National Institutes of Health and the Food and Drug Administration to improve the reporting of PROs and identify and validate existing PRO instruments. Currently, the quality of reporting and analysis of PROs in RCTs remains variable, leading to a CONSORT (Consolidated Standards of Reporting Trials) Statement extension outlining 5 items that are recommended for RCTs in which PROs are primary or secondary end points.<sup>1</sup>

The use of PROs in neurosurgical studies has thus far been limited and primarily exists in literature reporting patient outcomes following spine surgery. They have uncommonly been used in retrospective cohort studies to report adverse effects, resulting in disparities between the complication rates noted in these studies and large prospective trials, as the authors mention. Furthermore, PROs have rarely been reported in cerebrovascular and neurooncology studies, because the primary focus has typically been on variables such as perioperative complications,

recurrence or rebleeding rates, need for reoperation, and overall survival, among others. While important, these data points do not allow for the evaluation of the impact of our surgical and medical treatments on a patient's quality of life. With the increasing emphasis by regulators, policy-makers, and insurance companies on the need to incorporate PROs in therapy risk-benefit profiles, it is critical that we begin to evaluate these outcomes in future studies to not only ensure the continued reimbursement of our care, but also to assist with selecting the optimal treatment modality for patients with neurosurgical diseases.

<http://thejns.org/doi/abs/10.3171/2015.3.JNS142959>

## References

1. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al: Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* **309**:814–822, 2013
2. Campbell PG, Malone J, Yadla S, Chitale R, Nasser R, Maltenfort MG, et al: Comparison of ICD-9-based, retrospective, and prospective assessments of perioperative complications: assessment of accuracy in reporting. *J Neurosurg Spine* **14**:16–22, 2011
3. De Witt Hamer PC, Robles SG, Zwinderman AH, Duffau H, Berger MS: Impact of intraoperative stimulation brain mapping on glioma surgery outcome: a meta-analysis. *J Clin Oncol* **30**:2559–2565, 2012
4. Drewes C, Sagberg LM, Jakola AS, Gulati S, Solheim O: Morbidity after intracranial tumor surgery: sensitivity and specificity of retrospective review of medical records compared to patient-reported outcomes at 30 days. *J Neurosurg* [epub ahead of print August 7, 2015. DOI: 10.3171/2014.12.JNS142206]
5. Gnanasakthy A, Mordin M, Clark M, DeMuro C, Fehnel S, Copley-Merriman C: A review of patient-reported outcome labels in the United States: 2006 to 2010. *Value Health* **15**:437–442, 2012
6. Solheim O, Jakola AS, Unsgård G: Scientific alchemy and proposed gold standards of care. *World Neurosurg* **82**:e566–e567, 2014
7. Woodworth GF, Baird CJ, Garces-Ambrossi G, Tonascia J, Tamargo RJ: Inaccuracy of the administrative database: comparative analysis of two databases for the diagnosis and treatment of intracranial aneurysms. *Neurosurgery* **65**:251–257, 2009
8. 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

The authors report no conflict of interest.

## Response

**Ole Solheim, MD, PhD,<sup>1,2</sup> Asgeir Store Jakola, MD, PhD,<sup>1,4</sup> Sasha Gulati,<sup>1,2</sup> MD, PhD, Lisa Millgård Sagberg, MSc,<sup>1,2</sup> and Christina Drewes, MD<sup>3,5</sup>**

Departments of <sup>1</sup>Neurosurgery and <sup>3</sup>Anesthesiology, St. Olavs University Hospital; Departments of <sup>2</sup>Neuroscience and <sup>5</sup>Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway; and <sup>4</sup>Department of Neurosurgery, Gothenburg University Hospital, Gothenburg, Sweden

We would like to thank Dr. Babu and Dr. Sampson for their insightful editorial.

We very much agree that the low validity of retrospective review of medical records calls for better and validated measures of treatment-related morbidity in neurosurgery. Although we are throwing stones in our own glass house, we believe that it is time to address validity issues for common neurosurgical outcome variables. Lord Kelvin said: “If you cannot measure it, you cannot improve it.” Comparisons across neurosurgical outcome studies are usually invalid due to differences in assessment. The evidence is not improved by pooling the weaknesses in systematic reviews of heterogeneous studies of low quality and with ill-defined end points, as we have pointed out earlier.<sup>14</sup>

As neurosurgery has evolved into a more mature field, therapeutic quantum leaps are less frequent. Surgical results may nevertheless slowly improve over time due to adjustments of techniques, new technology, and improved patient selection. Because we can no longer expect to encounter enormous differences in treatment effects, sensitive and valid outcome measures are increasingly needed.

The practical, complex, and empirical nature of neurosurgery and the obstacles and limitations associated with RCTs still leave a major role for observational studies. Case series or cohort studies may serve as valuable audits for the conducting surgeons and institutions, equivalent to Phase IV drug trials (i.e., “post-market” surveillance of the procedures or tools). However, selection bias, assessment bias, and publication bias are common.<sup>16,17</sup> RCTs account for only 3% of neurosurgical studies found in MEDLINE, and only about half of such trials are of good quality.<sup>13</sup> While randomized controlled trials can assess efficacy (how interventions work under ideal settings in carefully selected patients), observational studies in unselected patients can assess effectiveness (how interventions work in non-ideal everyday settings).<sup>3</sup> However, in stark contrast to medical drug studies, effectiveness studies in surgery often report much better results than stringent efficacy studies. Thus, while RCTs in medicine are criticized for boosting treatment effects through careful patient selection and close follow-up, surgical RCTs may instead be accused of unacceptable poor treatment results of little relevance to expert surgeons, proving their excellence in retrospective case series (the classic “not in my hands” argument).

Given the steady predominance of case series in neurosurgery, a priority should be to develop and embrace standards of reporting in case series and other observational studies. While harmonization of assessments with, for example, the Good Clinical Practice Guidelines (<http://www.ich.org/home.html>) is enforced in drug studies, attempts at classifying common measures such as complications,<sup>7</sup> are not much embraced in neurosurgery. Validated neurological functional scales could allow for more valid and sensitive comparisons across studies than the more arbitrary or insensitive measures used today. Patient-reported outcomes at specified time points could be another step toward reducing assessment bias and emphasizing outcomes that are relevant to the patients themselves.<sup>6,12</sup> Assessing quality of life or quality of survival in brain operations that are essentially palliative in nature may, for example, be more relevant than extent of resection. Still,