Morbidity after intracranial tumor surgery: sensitivity and specificity of retrospective review of medical records compared with patient-reported outcomes at 30 days

Christina Drewes, MD,1,2 Lisa Millgård Sagberg, MSc,3–5 Asgeir Store Jakola, MD, PhD,1,5 Sasha Gulati, MD, PhD,3 and Ole Solheim, MD, PhD3,4,5

Departments of 1Anesthesiology and 2Neurosurgery and 3National Competence Centre for Ultrasound and Image-Guided Therapy, St. Olavs University Hospital; and Departments of 3Circulation and Medical Imaging and 4Neuroscience, Norwegian University of Science and Technology, Trondheim, Norway

OBJECT Published outcome reports in neurosurgical literature frequently rely on data from retrospective review of hospital records at discharge, but the sensitivity and specificity of retrospective assessments of surgical morbidity is not known. The aim of this study was to elucidate the sensitivity and specificity of retrospective assessment of morbidity after intracranial tumor surgery by comparing it to patient-reported outcomes at 30 days.

METHODS In 191 patients who underwent surgery for the treatment of intracranial tumors, we evaluated newly acquired neurological deficits within the motor, language, and cognitive domains. Traditional retrospective discharge data were collected by review of hospital records. Patient-reported data were obtained by structured phone interviews at 30 days after surgery. Data on perioperative medical and surgical complications were obtained from both hospital records and patient interviews conducted 30 days postoperatively.

RESULTS Sensitivity values for retrospective review of hospital records as compared with patient-reported outcomes were 0.52 for motor deficits, 0.4 for language deficits, and 0.07 for cognitive deficits. According to medical records, 158 patients were discharged with no new or worsened deficits, but only 117 (74%) of these patients confirmed this at 30 days after surgery. Specificity values were high (0.97–0.99), indicating that new deficits were unlikely to be found by retrospective review of hospital records at discharge when the patients did not report any at 30 days. Major perioperative complications were all identified through retrospective review of hospital records.

CONCLUSIONS Retrospective assessment of medical records at discharge from hospital may greatly underestimate the incidence of new neurological deficits after brain tumor surgery when compared with patient-reported outcomes after 30 days.


KEY WORDS brain tumor; acquired deficits; perioperative complications; patient-reported; outcome; oncology

No more than 10% of studies in 3 major neurosurgical journals in 2010 were prospective,13 and only around 3% of neurosurgical studies are randomized9—most of these comparing drugs, not surgical procedures. Even in neurosurgical randomized controlled trials (RCTs), blinded outcome investigation is seen in only 39% of studies.9 Consequently, retrospective case series still form the main basis for benchmarking risks and benefits of surgery and for comparing results between different techniques or centers. The lack of standardized outcome measures and follow-up times may potentially hamper comparisons across studies, centers, and populations. For softer and more subjective outcome measures, such as morbidity and complications, the tradition of unblinded retrospective assessment of surgeons’ own data may be especially prone to bias. In a large meta-analysis of studies on glioma surgery,3 the risk of acquired deficits was 20.1% in the 2 RCTs, whereas in 88 observational, mainly retrospective case series, deficits were observed in only 4.5% of patients.10 Also in publications on man-
agement of unruptured intracranial aneurysms, studies of high methodological quality present higher risks of unfavorable outcome than low-quality studies. The question remains whether reviewing medical records in retrospect provides a realistic impression of surgical morbidity or merely shows the tip of the iceberg.

In a study population of 191 patients who underwent planned resection of intracranial tumors, we assessed postoperative morbidity based on retrospective review of medical records compared with morbidity reported by the patients themselves.

Methods

In this registry-based study, we included 191 (46.8%) of 409 adult patients (age ≥ 18 years) who underwent craniotomy for the treatment of intracranial tumors at Trondheim University Hospital, Norway, in the period from September 9, 2011, to October 10, 2013. Our department serves a geographically defined catchment region. Only patients with confirmed intracranial tumors who gave their written informed consent and who later provided follow-up data were included in the study.

A flowchart of the inclusion process, with reasons for noninclusion, is shown in Fig. 1.

Data collection was approved by the Regional Ethical Committee for Health Region Middle Norway and adhered to the guidelines of the Helsinki Declaration. Newly acquired or worsened motor deficits, language deficits, and cognitive deficits were included as outcome measures. Traditional (i.e., retrospective) outcome data were obtained by thorough review of available documentation from hospital medical records (electronic). Referral documentation (from other hospitals or general practitioners), admission notes (by neurosurgical registrars), and surgery reports (by operating neurosurgeons) were reviewed to record preoperative function/deficits. Postoperative documentation, such as notes and reports by neurosurgical doctors, neurologists, and/or oncologists, including discharge summaries, were closely reviewed to assess function at discharge. Prospective data on patient-reported new or worsened deficits were collected by one of 2 research nurses in a structured phone interview 30 days postoperatively (median 30 days, range 23–52 days). Proxies (close relatives) were consulted in cases involving patients with severe communication problems or cognitive deficits. The patients were asked for “limb weakness” or “palsy” that was new or had worsened after surgery. If the answer was “yes,” the nature of the new deficit (arm or leg weakness/hemiparesis or -paralysis) was recorded. For recording cognitive worsening or language deficits, the patients were asked for “new/worsened cognitive/memory problems (yes/no)” after surgery and for “new/worsened speech problems (yes/no)” after surgery. In cases in which patients expressed that they were unsure about the existence of new or worsened deficits, the response was classified as “no.”

Data on acquired neurological deficits as registered at discharge and new neurological deficits as reported by the patients themselves at 30 days postsurgery were dichotomized (new or worsened deficits, yes/no) and compared for sensitivity and specificity. Additionally, we analyzed postoperative complications registered in the hospital medical records within the first 30 days after surgery and compared the results on a case by case basis with complications as reported by the patients at 30 days.

Statistical Analyses

All statistical analyses were performed using SPSS Statistics 21.0. The statistical significance level in tests of
normality was set to $p < 0.05$. Q-Q plots were used to test for normal distribution of data. Descriptive statistics are presented. Sensitivity and specificity were calculated from $2 \times 2$ contingency tables.

Sensitivity and specificity are measures of test properties, and, in this context, patient-reported outcomes were set as the “true value,” while outcome data from the hospital records served as the clinical test with either positive or negative results.

**Results**

As seen in Table 1, the mean age in our patient cohort was 54 years, and 48% of the patients were female. Significant preoperative comorbidity as indicated by American Society of Anesthesiologists (ASA) Class 3 and Charlson Comorbidity Index $> 5$ was seen in 24% and 8% of patients, respectively. Prior to surgery, 171 patients (90%) were able to care for themselves, as indicated by Karnofsky Performance Status scores $\geq 70$.

**No New Deficits**

New or worsened deficits as documented in medical records compared with patient reports are summarized in Table 2. According to medical records, 158 patients were discharged with no new or worsened deficits, but only 117 (74%) of these patients confirmed this.

**Motor Deficits**

Medical records documented 22 patients with new or worsened motor deficits after surgery. In 16 of the patients who reported new or worsened motor deficits at follow-up, there was no corresponding documentation in hospital medical records. In 5 patients for whom new or worsened motor deficits were registered in medical records at discharge, the patients denied the development or worsening of the deficit during the follow-up interview 30 days postoperatively. The specificity of medical record documentation compared with the patients’ experience was 0.97, but sensitivity was only 0.52.

**Language Deficits**

Medical records documented new or worsened language deficits in 13 patients after surgery but failed to do so in 18 patients who themselves reported new deficits at follow-up. One patient denied any history of language deficits at follow-up despite documentation of new language deficit in the medical record. For language deficits, the specificity of retrospective review of medical records as compared with patient-reported outcomes was 0.99, but sensitivity was only 0.52.

**Cognitive Deficits**

Medical records documented new or worsened cognitive deficits in 5 patients after surgery, but failed to report new cognitive problems in 29 patients who reported new or worsened cognitive deficits at follow-up. Three patients denied any new or worsened cognitive deficits at follow-up despite documentation of new or worsened cognitive deficits in the medical records. For cognitive deficits the specificity of retrospective review of medical records as compared with patient-reported outcomes was 0.98, but sensitivity was only 0.07.

For all 3 evaluated deficit categories, Fig. 2 depicts the difference in reporting between hospital records at discharge and patient-reported outcome at 30 days.

**Perioperative Complications**

Comparison of hospital records and patient-reported data showed that serious complications such as postoperative intracranial hematomas, perioperative stroke, hydrocephalus, meningitis, and deep surgical-site infections, were all documented in hospital medical records within the first 30 days (Table 3). Two superficial wound infections were reported by the patients at 30 days and not documented in the hospital records as occurring within the first 30 days after surgery, but were treated and documented in the hospital records more than 30 days postoperatively. Six urinary tract infections were reported by patients at 30 days but were not registered in the hospital records.

**Discussion**

Retrospective review of hospital records at discharge is a traditional and widely used method of assessing ac-
required deficits in the neurosurgical literature. In the present study, we compared this to patient-reported outcomes at 30 days. We found that assessments based on retrospective reviews of medical records may greatly underestimate postoperative neurological deficits. Sensitivity of retrospectively obtained data from hospital records was low when compared with the patients’ perspectives at 30 days after surgery. Especially cognitive problems seemed to be largely underestimated by review of discharge data. This is not surprising, because subtle, yet important changes in cognitive function might not even occur to the patients themselves during the first days after surgery, and consequently, not to the surgeon either. The same may apply to language deficits, although to a considerably smaller degree, with a sensitivity value of 0.40 for the retrospective review of hospital medical records. The discordance rate of 11% between the surgeons’ reports on new motor deficits compared with the patients’ does, at first glance, seem acceptable. Nevertheless, the sensitivity of the surgeon’s assessment of new motor deficits was no more than 0.52. For this category of deficits as well, it has to be considered that problems with fine motor skills such as writing, crafting, or other manual work may not always be detected by a customary neurological assessment at the end of a hospital stay. The relatively high specificity values we obtained for retrospective review of medical records indicate that new deficits were unlikely to be documented in the hospital records at discharge when the patients did not report any at 30 days.

In gastrointestinal surgery, a systematic review revealed that anastomosis leak was defined in 54 different ways in the literature. Similar studies are lacking in neurosurgery, but for softer outcome measures such as surgical morbidity and even extent of resection, there is no consensus on outcome classification. Therefore, it is no surprise that the heterogeneity of the systematic review on intraoperative mapping by De Witt Hamer and colleagues was almost 100% for early deficits and 87% for late deficits. When outcomes are measured systematically and prospectively, as for example in the Glioma Outcome Project and the 5-ALA (5-aminolevulinic acid) study, surgical morbidity is strikingly much higher than commonly reported in retrospective case series. Together with the known publication bias, the heterogeneity and, as seen, the insensitivity of retrospective outcome assessment is clearly problematic for benchmarking results, comparing results, and for meta-analyses in neurosurgery.

The sensitivity and specificity of retrospective reviews of medical records is likely to depend considerably on local clinical routines and time of follow-up. Some authors define early and late outcome assessment, but there are large variations in timing, assessing profession (e.g., surgeon, neuropsychologist), assessment methods (neurological examination, neuropsychology examination, rating scales), and assessed domains. Regardless, there is a need for consensus on how and when to assess surgical morbidity following brain tumor surgery. Landriel Ibañez and colleagues have suggested a classification system of surgical and medical complications (including new deficits) in neurosurgery, but unfortunately, this system has not been widely embraced yet. Whether patient-reported measures are an optimal solution can be debated. Advantages of

<table>
<thead>
<tr>
<th>Description</th>
<th>Motor</th>
<th>Language</th>
<th>Cognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement b/wn hospital records &amp; pt reports</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No new deficit</td>
<td>153</td>
<td>160</td>
<td>157</td>
</tr>
<tr>
<td>New deficit(s)</td>
<td>17</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Sum, agreement</td>
<td>170 (89%)</td>
<td>172 (91%)</td>
<td>159 (83%)</td>
</tr>
<tr>
<td>Discordance b/wn hospital records &amp; pt reports</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficit reported by patients only</td>
<td>16</td>
<td>18</td>
<td>29</td>
</tr>
<tr>
<td>Deficit reported by clinicians only</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Sum, discordance</td>
<td>21 (11%)</td>
<td>19 (9%)</td>
<td>32 (17%)</td>
</tr>
<tr>
<td>Sensitivity of retrospective review of hospital records</td>
<td>0.52</td>
<td>0.40</td>
<td>0.07</td>
</tr>
<tr>
<td>Specificity of retrospective review of hospital records</td>
<td>0.97</td>
<td>0.99</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Pt = patient.

FIG. 2. New or worsened deficit documented in discharge hospital records (HR) compared with patient-reported outcome at 30 days (PRO).
standardized and validated patient-reported measures include prospective data collection and reduced impact of diverging local traditions for outcome assessment. Another advantage of patient-reported data is limiting personal interest bias that might affect results when surgeons assess the outcome of their own operations.

In our study population, major medical and surgical complications (excluding new deficits) were all identified through retrospective review of hospital records within the first 30 days, almost two-thirds already at discharge. A few urinary tract infections were reported by patients only. Retrospective review of hospital records at 30 days may give a realistic impression of the complication rate.

The nonstandardized method and timing of assessment of neurological functions following brain tumor surgery is clearly problematic. Neurological function is not stable after surgery, and subtle deficits can be difficult to detect during a routine gross examination. We believe that a scheduled, detailed, and standardized neurological evaluation at 30 days after surgery might have yielded higher sensitivity values than the present retrospective review of medical records at discharge. Still, high sensitivity compared with patient-reported results is not necessarily the most crucial aspect of an outcome measure. To avoid confusion and to facilitate comparison of outcomes, reliability and validity are by far most important. Unfortunately, validated measures of neurosurgical morbidity are lacking for intracranial surgery.

Strengths/Weaknesses

In the present study, the population-based referral is a strength that ensures a well-defined and representative cohort. However, some patients did not consent to participate, possibly due to a wish to deal with their severe diagnosis without having to meet the additional demands from participation in a study. Some patients were lost to follow-up, mainly due to administrative errors and, in 3 cases, death. There may have been inconsistencies in reporting from cognitively impaired patients. Also, external validity of our findings depends on local clinical routines of assessing and documenting surgical results in medical records. In our hospital, postoperative outcome assessment at discharge is documented based on a nonstandardized clinical neurological assessment, without the use of a classification system. It has to be considered that a defined classification system or other specialists like neuropsychologists or speech therapists might have unveiled a greater part of the less-obvious deficits already at the time of discharge. Nevertheless, we believe that our clinical practice is in accordance with data published in many, if not most neurosurgical case series, which enhances the relevance of our results.

Conclusions

Retrospective assessment of medical records at discharge from hospital may greatly underestimate the incidence of new neurological deficits after brain tumor surgery when compared with patient-reported outcomes after 30 days. Although routines for when and how to assess surgical morbidity vary between institutions and publications, we believe our findings are relevant for a disconcertingly large proportion of the neurosurgical outcome literature.

References


**Disclosure**

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

**Author Contributions**

Conception and design: all authors. Acquisition of data: Drewes, Sagberg, Jakola, Solheim. Analysis and interpretation of data: Drewes. Drafting the article: Drewes, Solheim. Critically revising the article: all authors.

**Correspondence**

Christina Drewes, Norwegian University of Technology and Science, Department of Circulation and Medical Imaging, Postboks 8905, Trondheim 7491, Norway. email: christina.drewes@ntnu.no.