Prior to general use of a new procedure there is widespread belief in its efficacy. In surgery, arrival at this point, essentially acceptance from the relevant community, remains an ambiguous process. Unlike drug development, surgical procedures do not follow a standardized developmental pathway or always arrive at clear approval from regulatory agencies. Currently, surgeons have considerable freedom in using untested procedures and it is mostly up to the surgical community to police itself. This results in a variety of rationalizations for use, including communications of personal experience or historical usage, in addition to clinical studies. Yet recently calls have intensified from payers, regulators, and many in the medical community to limit the definition of efficacy to high-level evidence, chiefly that from randomized controlled trials (RCTs). For better or worse, a major consequence of such adoption will be alterations in the way surgeons investigate new procedures.

**OBJECTIVE** Patients, practitioners, payers, and regulators are advocating for reform in how medical advances are evaluated. Because surgery does not adhere to a standardized developmental pathway, how the medical community accepts a procedure remains unclear. The authors developed a new model, using publication data and patterns, that quantifies this process. Using this technique, the authors identified common archetypes and influences on neurosurgical progress from idea inception to acceptance.

**METHODS** Seven neurosurgical procedures developed in the past 15–25 years were used as developmental case studies (endovascular coil, deep brain stimulation, vagus nerve stimulation, 1,3-bis(2-chloroethyl))-l-nitrosourea wafer, and 3 radiosurgery procedures), and the literature on each topic was evaluated. A new metric the authors termed “progressive scholarly acceptance” (PSA) was used as an end point for community acceptance. PSA was reached when the number of investigations that refine or improve a procedure eclipsed the total number of reports assessing initial efficacy. Report characteristics, including the number of patients studied, study design, and number of authoring groups from the first report to the point of PSA, were assessed.

**RESULTS** Publication data implicated factors that had an outsized influence on acceptance. First, procedural accessibility to investigators was found to influence the number of reports, number of patients studied, and number of authoring groups contributing. Barriers to accessibility included target disease rarity, regulatory restrictions, and cost. Second, the ease or difficulty in applying a randomized controlled trial had an impact on study design. Based on these 2 factors, 3 developmental archetypes were characterized to generally describe the development of surgery.

**CONCLUSIONS** Common surgical development archetypes can be described based on factors that impact investigative methods, data accumulation, and ultimate acceptance by society. The approach and proposed terminologies in this report could inform future procedural development as well as any attempts to regulate surgical innovation.
Effective reform requires critical evaluation of how innovations currently develop from conception to general acceptance. The goal of this study is to assess the publication patterns, such as study volume, design, and dissemination, of several key neurosurgical procedures to identify common archetypes and influences in the development of neurosurgery.

Methods
Procedure Selection
To investigate patterns of innovation in neurosurgery, 7 procedures were chosen to act as case studies of development. The development of each procedure was individually characterized, and then the results from all were compared for similarities, differences, and patterns.

Selected procedures must have been conceived long enough ago (15–25 years) that initial questions of efficacy had been answered and hindsight could be used to interpret developmental patterns. Procedures were selected across neurosurgical subfields and from a cluster in 1 field (stereotactic radiosurgery [SRS]) to compare similarities across neurosurgical subfields and from a cluster in 1 field. The development of each procedure was individually characterized, and then the results from all were compared for similarities, differences, and patterns.

Six were selected as examples of procedures that ultimately demonstrated efficacy: 1) radiosurgery for trigeminal neuralgia (TN), 2) radiosurgery for meningioma, 3) radiosurgery for brain metastasis, 4) deep brain stimulation (DBS) for essential tremor, 5) endovascular coiling of aneurysms, and 6) 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU) polymer insertion during malignant glioma surgery. A seventh was selected as an example of a procedure with initial promise that proved ineffective: vagus nerve stimulation (VNS) for treatment-resistant depression.

Defining Community Acceptance
Identifying a standard end point of when a procedure was generally accepted by the community was necessary for this analysis, yet this was difficult due to the nonstandard nature of surgical innovation. We developed a new metric in our analyses called progressive scholarly acceptance (PSA) as an end point for community acceptance. The theory behind PSA and techniques for modeling it are explained in more detail in Part 1 of this report. In brief, the idea of PSA stems from the common view of surgical innovators as “tinkerers” who continually refine procedures. This view is historically supported, and surgical innovation can be viewed as a series of broad innovations, followed by periods of refinement. Key to the PSA theory is the recognition that prior to improvement comes acceptance of the broad, initial innovation. That is, before an investigator begins to study a refinement, he or she believes in the procedure they are building on.

As a procedure demonstrates efficacy and its use increases, more authors will begin to study and publish their own refinements of the innovation. At some point, the number of refining studies will become larger than the number of efficacy studies. This is an indication of community acceptance; the majority of the community has then progressed beyond the initial question of efficacy into refinement. This is the point designated as PSA.

Data Collection
To measure PSA, each procedure was searched in PubMed (www.ncbi.nlm.nih.gov/pubmed/) with essential key words, including historical names. Additional filters were applied to limit review to human studies written in English. All research studies were reviewed in chronological order. Each paper was characterized as either an “initial investigation,” which addressed the question of broad efficacy, or as a “refining study.” Possible refining studies could be investigating a modified procedure method, the procedure in a narrower population, or use of the procedure in a particular disease subtype. More details on differentiating initial from refining studies are presented in Part 1 of this report.

In addition to publication dates and defining the studies as initial or refining, other report characteristics were recorded to investigate development patterns. The number of patients studied (not including control subjects), the authors and main site of investigation, the study design, and whether the study was conducted in the United States or abroad were all noted.

PSA Analysis
The first step in our analysis was to identify the PSA point—the end point used for community acceptance. We used 2 PSA models: annual and compounding. In the annual analysis, the number of initial investigations and number of refining studies per year were plotted against time as line graphs (Figs. 1A–7A). With this model, the annual number of studies can be seen, and fluctuations each year due to many potential influences can be observed. But that fluidity, which is valuable to see the impact of a certain factor on the number of publications, makes it less useful for confidently identifying acceptance. Refinement could become more common than initial investigation for 1 year without being sustained. A more conservative approach for recognizing acceptance comes from the compounding model. It is similar to the annual model—total studies are plotted against time, with initial and refining studies having their own line graphs—except the total number of initial and refining publications is compounded each year (Figs. 1B–7B). This means that where the refining line surpasses the initial line, if it does so, is the point when the total number of reports improving a procedure has surpassed the total number initially studying its efficacy. For this to occur, the number of refining investigations would have to be higher than initial efficacy studies for some time, indicating sustained refinement and slower growth in studying efficacy. This implies the community has mostly embraced the procedure, and subsequent improvement has become of greater interest. This is the PSA point, which is the end point for our analyses.

Additional Publication Characteristics
In addition to the number of publications, several other report characteristics were investigated to better understand patterns of development in neurosurgery. For all data, the starting point was the first modern publication on the topic and the end point was the year where PSA was reached. Characteristics were analyzed for both initial
and refining studies, though extra emphasis was placed on the outcomes of initial studies, since these are the investigations that test initial efficacy and should indicate the burden of evidence required for community acceptance.

The total number of patients studied (control subjects not included) was recorded. This was only an estimate of total patients studied, as some studies may have included follow-up data or more specific investigations of the same patients in later studies (resulting in double counting). The average number of patients per study was calculated for each procedure. The design of each publication was classified as a case series, cohort study, case–control study,
cross-sectional study, or RCT. Classification of observational studies was based on the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist,30 with additional specificity on differentiating case series and cohorts based on Dekkers et al.9 Cohort studies provide absolute risk and outcome predictions based on clinical data.9

To better understand influence from authors and to observe dissemination, each publication was assigned to an “authoring group.” Reports were considered in the same authoring group if the studies shared 2 or more coauthors or had the same corresponding or first author. In the case of large multisite investigations with multiple authoring groups, the report was considered from a single group credited to the corresponding author. It is important to note that the goal of the authoring group metric was to view how the procedure disseminated to new investigators or investigative groups, not the total number of authors studying the device.

One difficult-to-measure metric considered for each procedure was “objective efficacy.” For each procedure, this was the point when clinical studies had conclusively demonstrated procedural efficacy (or lack thereof). This is important because community acceptance of a procedure may not always correspond to when efficacy was objectively demonstrated. Identifying potential causes of this discordance could be beneficial for preventing improper use. Unfortunately, because of the nonstandard nature of neurosurgical development, there is no consistent method to identify objective efficacy across all procedures. So to compare how PSA matches up with efficacy, several different methods, and some combination of them, were used to attempt to identify a point of objective efficacy.

One difficult-to-measure metric considered for each procedure was “objective efficacy.” For each procedure, this was the point when clinical studies had conclusively demonstrated procedural efficacy (or lack thereof). This is important because community acceptance of a procedure may not always correspond to when efficacy was objectively demonstrated. Identifying potential causes of this discordance could be beneficial for preventing improper use. Unfortunately, because of the nonstandard nature of neurosurgical development, there is no consistent method to identify objective efficacy across all procedures. So to compare how PSA matches up with efficacy, several different methods, and some combination of them, were used to attempt to identify a point of objective efficacy.

**Fig. 5. PSA for endovascular coiling of aneurysms, Years 1991–1997.** A: Annual publication of initial and refining reports. B: The number of annual initial (gray) and refining (black) reports compounded each year. C: The number of reports from each authoring group (separate shade for each group).

**Fig. 6. PSA for BCNU polymer insertion at malignant glioma, Years 1991–2003.** A: Annual publication of initial and refining reports. B: The number of annual initial (gray) and refining (black) reports compounded each year. C: The number of reports from each authoring group (separate shade for each group).

**Fig. 7. PSA for vagus nerve stimulation for treatment-resistant depression, Years 2000–2010.** A: Annual publication of initial and refining reports. B: The number of annual initial (gray) and refining (black) reports compounded each year. C: The number of reports from each authoring group (separate shade for each group).
Table 1. Path to acceptance for neurosurgical indications

<table>
<thead>
<tr>
<th>Procedure</th>
<th>No. of Initial Studies</th>
<th>No. of Initial Patients</th>
<th>No. of RCTs (%)</th>
<th>No. of Cohort Studies (%)</th>
<th>No. of Case Series (%)</th>
<th>No. of Initial Authoring Groups</th>
<th>Average No. of Patients/Initial Study</th>
<th>Years to PSA</th>
<th>Years to Objective Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiosurgery for TN</td>
<td>16</td>
<td>1250</td>
<td>0 (0)</td>
<td>16 (100)</td>
<td>0 (0)</td>
<td>9</td>
<td>78.1</td>
<td>10</td>
<td>10–11</td>
</tr>
<tr>
<td>Radiosurgery for meningioma</td>
<td>24</td>
<td>1247</td>
<td>0 (0)</td>
<td>22 (92)</td>
<td>2 (8)</td>
<td>17</td>
<td>52.0</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Radiosurgery for brain metastasis</td>
<td>25</td>
<td>1753</td>
<td>0 (0)</td>
<td>25 (100)</td>
<td>0 (0)</td>
<td>17</td>
<td>70.1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>DBS for essential tremor</td>
<td>13</td>
<td>229</td>
<td>0 (0)</td>
<td>13 (100)</td>
<td>0 (0)</td>
<td>10</td>
<td>17.6</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Endovascular coiling aneurysms</td>
<td>10</td>
<td>317</td>
<td>0 (0)</td>
<td>6 (60)</td>
<td>4 (40)</td>
<td>7</td>
<td>31.7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>BCNU polymer wafers for glioma</td>
<td>4</td>
<td>267</td>
<td>3 (75)</td>
<td>1 (25)</td>
<td>0 (0)</td>
<td>3</td>
<td>66.8</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>VNS for depression</td>
<td>4</td>
<td>131</td>
<td>0 (0)</td>
<td>4 (100)</td>
<td>0 (0)</td>
<td>2</td>
<td>32.8</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Objective efficacy was reached. For 2 procedures studied (SRS for meningiomas and brain metastases), a single year to designate as the point of objective efficacy could not be confidently identified because of the gradual demonstration of efficacy of the procedures.

Interpreting Patterns and Statistical Analysis

After completion of data collection and metric computation, data were reviewed for patterns. When comparing characteristics of 2 groups of procedures, 2-sample, unpaired t-tests were used to test difference. Results were reviewed and patterns considered with knowledge of the procedures’ backgrounds, such as regulatory agency rulings and characteristics of the studied disease. This was a limitation of this investigation, as any retrospective analysis inherently comes with risk of bias in interpretation.

Results

Stereotactic Radiosurgery for TN

The first report of modern techniques for treating TN, also known as tic douloureux, with SRS was in 1993 by Rand et al. While using radiosurgery for treating TN was also known as tic douloureux, with SRS was in 1993 by Rand and colleagues. A larger, multicenter report that included 50 patients was reported in 1996. Over the next decade, the total number of initial reports grew at a relatively steady rate as new authors began investigating SRS for TN (Fig. 1). By 2000, the majority of authors were publishing refining investigations, such as using higher doses of radiation, or specifically investigating patients with tumor-related TN. By 2002, the total number of refining studies (n = 20) overtook the total number of initial studies (n = 16). So 2002 is considered the point of PSA, and characteristics for SRS for TN were evaluated from papers published from 1993 through 2002.

Stereotactic Radiosurgery for Brain Metastases

Publication of SRS for a single brain metastasis began in 1987 with a report by Sturm and colleagues. There was some lull in publication as researchers began to apply the new technique and gather data, and then a large influx in publications started around 1990. Researchers began studying SRS for single and multiple brain metastases in 1992, though we did not consider the application to mul-
multiple metastases as refining in our analysis because soon after the onset of studying multiple metastases, the field largely transitioned to using both solitary and multiple metastases for both continued efficacy and refining studies (e.g., dosage, different tumor types, and so forth).

PSA was reached in 1998, 12 years after the first publication on the topic, with 25 initial studies by 17 authoring groups and 28 refining investigations. Of the initial studies, all were cohort studies, with an average of 70.1 patients per study and a total of 1753 patients studied (Table 1).

Similarly to SRS for meningiomas, no single year could be accurately identified as the point of objective efficacy for SRS for brain metastases, as payers gradually accepted the procedure and there were no national coverage determinations.

Deep Brain Stimulation for Essential Tremor

The first study to assess DBS to treat essential tremor (ET) was published in The Lancet in 1991 and conducted by Benabid et al. The rate of publication remained steady at 1 to 2 articles per year for the next 10 years. In July 1997, the FDA approved unilateral thalamic stimulation for suppression of tremor in ET and Parkinson’s disease. This was the measure used to designate objective efficacy, which took 7 years to reach.

Also in 1997, the rate of refinement studies increased considerably, and PSA was reached by 1999. Time to PSA was 9 years, with 13 initial and 14 refining investigations. All 13 initial investigations (100%) were classified as cohort studies. While many of the studies used blinded evaluations by randomly turning off the stimulating de-
vice hours before clinical follow-up several months after the operation, these were not considered randomized controlled trials, as all patients received intervention with full effect of stimulation immediately following the operation.

A total of 229 patients were enrolled in initial studies, with an average of 17.6 patients per study. Ten separate authoring groups contributed to the 13 initial studies.

**Endovascular Coiling for Cerebral Aneurysms**

We specifically reviewed the development of Guglielmi detachable coils (GDCs) to treat cerebral aneurysms. The first preliminary clinical experience of GDCs was published by inventors Guglielmi et al. in 1991. Initial inquiry continued in 1992, then decreased in 1993, and no efficacy studies were published in 1994 (Fig. 5A). There was a rebound of initial inquiry in 1995 that remained steady for the remainder of the years that data were collected (through 1997).

As initial inquiry decreased in 1993 and 1994, refining investigations increased significantly. A higher level of refinement was sustained from 1994 onward, and PSA was reached in 1995, after 5 years, with 11 initial studies and 11 refining studies (Table 2). Of those 11 initial studies, 7 (64%) were cohort studies and 4 (36%) were case series (Table 1). Eight different authoring groups contributed to the 11 initial studies. Prior to PSA, 324 patients were studied, with an average of 29.5 patients per study.

The FDA approved GDCs in 1995 (FDA Premarket Notification K951256; https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K951256). This approval was based on early results from a large, 8-center cohort trial with a cohort of 403 patients, which was not published until 1997 (after PSA was reached, and so not included in these analyses). The point of objective efficacy for coiling of cerebral aneurysms was considered to be 1995, 5 years after first publication.

**BCNU Polymer Wafers for Malignant Gliomas**

The first investigation of Gliadel (BCNU polymer wafers; Arbor Pharmaceuticals) was a cohort study of 21 patients conducted by the inventors, Brem et al., and was published in 1991. Four years later, the same authoring group published a larger, multicenter RCT. The FDA approved the treatment for recurrent malignant gliomas 1 year later in 1996 (FDA Highlights of Prescribing Information Reference ID 3358686; http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K951256). This approval was based on early results from a large, 8-center trial with a cohort of 403 patients, which was not published until 1997 (after PSA was reached, and so not included in these analyses). The point of objective efficacy for coiling of cerebral aneurysms was considered to be 1995, 5 years after first publication.

The point of objective efficacy was designated as the year that the FDA approved the drug, 6 years from the first report.

**Vagus Nerve Stimulation for Treatment-Resistant Depression**

The first trial intentionally testing (not simply noting an incidental observation) whether VNS would benefit patients with treatment-resistant depression was a multicenter cohort trial in 2000 conducted by Rush et al. This authoring group continued to study VNS for depression and several additional authoring groups began to publish both efficacy and refining studies. By 2004, the total number of refining studies outnumbered the initial studies, indicating PSA (Fig. 7B).

At the point of PSA, 5 years after study of the procedure began, 4 initial studies and 5 refining studies were published (Table 2). Of the 4 initial studies, all were cohort studies and 2 authoring groups had contributed (though the first comprised 4 study sites). A total of 131 patients were enrolled, with an average of 32.8 patients per study (Table 1).

In 2005, a large, 21-center RCT assessing VNS for depression was published (it is not included in our analyses, since it arrived after PSA). The RCT found no evidence of efficacy for VNS for resistant depression. Still, in the same year, the FDA approved the treatment. Then, in 2007, the CMS issued a National Coverage Analysis concluding, based on the 2005 RCT, there was sufficient evidence that the treatment was not an effective treatment for depression. Since both of these regulatory analyses were based on studies published in 2005 or earlier, the point of objective efficacy, when the effectiveness of a procedure can be discerned, was considered to be 2005—6 years of study since the procedure’s introduction.

**Statistical Analysis**

To investigate the impact of how regulatory restrictions affect developmental characteristics, the data from procedures that required FDA approval and those that did not were compared. The 3 procedures that did not require additional FDA approval (SRS for TN, meningiomas, and brain metastases, as Gamma knife was broadly approved for creation of brain lesions by the FDA in 1989) were sorted into 1 group and the procedures that did require FDA approval (DBS for essential tremor, GDCs for cerebral aneurysms, BCNU polymer wafers for malignant gliomas, and VNS for depression) were sorted into another group. Two-sample, unpaired t-tests were used to compare characteristics from each group. For the group that did not require FDA approval, there was a significantly larger number of total initial studies conducted (p = 0.0115), more initial patients studied (p = 0.0005), and a larger number of total authoring groups that conducted initial studies (p = 0.0368), but there was no significant difference found for the average number of patients per initial study (p = 0.088).

**Discussion**

Many payers, regulators, practitioners, and patients are...
advocating for reform in the way that medical advances are evaluated. Increasingly, efficacy is limited to drugs and procedures that have demonstrated effectiveness in RCTs. While increasing evaluative stringency is well meaning, characteristics inherent to surgical procedures may make RCTs an imperfect tool for their valuation. Before the methodology for investigating surgical procedures can be improved, the current methods and factors that influence development need to be better understood. The goal of this study was to assess the evaluation and publication characteristics of an array of neurosurgical procedures, to identify common archetypes and potential influences of development.

Surgical procedures are not all the same. Each procedural characteristic or the disease it addresses has the potential to impact development. This makes the study of surgical development a challenge and no single model will be able to accurately describe the way all procedures move from conception to acceptance. Though, from reviewing the results of the 7 neurosurgical procedures studied here, there do appear to be several factors that have an outsized influence on the number of reports published, the number of patients studied, the number of authoring groups contributing, and the time to community acceptance.

Accessibility of the Procedure to Researchers

The first major factor is the accessibility of the procedure to researchers. If researchers can more easily gain access to and study a procedure, it makes sense that more authoring groups will publish more studies on the topic and more patients will be studied. Our examples confirm this and demonstrate several key barriers to accessibility.

One uncontrollable barrier to access that could limit investigation is disease rarity or the difficulty in administration of a proposed treatment. If a disease is less common or a procedure particularly difficult to perform, it is likely that fewer patients will be recruited per study. This barrier is best viewed as impacting the average number of patients per study, since the total number of patients studied could be equal between a rare illness and common illness if the rare disease is studied long enough and the common disease studied only briefly.

The impact of this barrier can be observed with the SRS examples. The tools required for all 3 SRS procedures, treating TN, brain metastases, and meningiomas, are the same, but the disease characteristics alter accessibility. Meningiomas are comparatively rare tumors that progress slowly, while brain metastases occur more commonly and their full trajectory can be observed in a shorter period of time. This difference could explain why the average number of meningioma patients per study (n = 52.0) was smaller than the average number of brain metastases patients per study (n = 70.1). The average number of meningioma patients per paper was also smaller than the average number of TN patients (n = 78.1) per study, even though the total number of patients studied are almost identical (n = 1247 and 1250, respectively), since there were more total investigations of meningiomas.

The Effect of Regulatory Restrictions on Accessibility

Another barrier to accessibility is regulatory restric-

For the procedures studied, there was a significant difference in publication characteristics between the procedures requiring FDA approval and those that did not. Those that did not require FDA approval had higher total numbers of initial, or broad efficacy, studies (p = 0.0115); higher total numbers of patients in initial investigations (p = 0.0005); and a larger number of authoring groups investigating efficacy (p = 0.0368) prior to community acceptance. And while rarity of disease or ease of administration could be a factor in these differences, there was no significant difference in the average number of patients per initial study (p = 0.088) between the 2 groups. This indicates that large enrollment from more common diseases or easier-to-administer procedures do not fully explain the differences in the total number of patients, and number of studies or authoring groups. These results are best explained by a decrease in researcher accessibility to procedures for which FDA approval is still pending.

It is important to emphasize that accessibility is not a categorical all-or-nothing variable. For example, while access to Gliadel was restrictive and only the investigators who participated in the main trial seeking FDA approval could study the treatment, the story of DBS for ET is more nuanced. Devices for DBS were restricted for ET in the United States until FDA approval in 1996 (FDA Highlights of Prescribing Information Reference ID 3358686; http://www.accessdata.fda.gov/drugsatfda_docs/label/2013/020637s026lbl.pdf), but they were given CE marking in Europe in 1993, which can be a less stringent European approval process for devices. This increased accessibility in Europe, and, in fact, all investigations prior to 1996 were conducted in Europe. So DBS for ET could be considered in the middle of the spectrum of researcher accessibility: accessible in Europe and heavily restricted in the US prior to FDA approval. This could explain why DBS for ET and GDCs (CE approval in 1992) had more authoring groups investigating efficacy than other procedures that required FDA approval.

Another probable barrier to accessibility is cost. More-expensive procedures require more funding and likely inhibit access to researchers. Cost was not evaluated for these procedures, though its influence on restricting access to new procedural investigation seems likely and demands further inquiry.

Barriers to Randomized Controlled Trials

The second most significant factor besides accessibility is the difficulty in conducting an RCT for that procedure. There are many potential barriers to using RCTs to investigate the efficacy of a surgical procedure. They include funding for an expensive clinical study if the procedure is nonproprietary, ethical questions of sham surgeries or of equipoise, low recruitment rates, and practical questions of blinding and controlling,1,15,26
The difficulty in applying an RCT is the major factor impacting the type of study designs used to investigate procedures. This can be seen in our examples. Gliadel was a proprietary treatment that had a randomization process similar to most new drug treatments. After tumor resection, polymer discs, with the active drug or without, were implanted onto resection cavity surfaces. Neither investigators, monitors, nor patients knew whether they received the active drug or placebo. Because of how well it lent itself to RCT application, the majority (75%) of trials used to evaluate Gliadel were RCTs. This is in contrast to the development of SRS for brain metastases. For the radiosurgery procedure, all the initial studies were cohort studies. Several barriers to RCTs in this setting, including difficulty in blinding, questions of equipoise, and limited funding, since the procedure was nonproprietary, could explain why, to date, there has never been a meaningfully conducted randomized trial of surgical resection using radiosurgery.

As with researcher accessibility, the difficulty in applying RCTs can be seen across a spectrum, which depends on how insurmountable the barriers to RCTs are for the particular procedure. For example, while Gliadel was the only example in this analysis for which an RCT was conducted prior to PSA, authors later managed to test efficacy using an RCT of several other procedures studied (DBS for ET, GDC for aneurysms, and VNS for depression). So for these procedures, application of an RCT was possible, but was not a first-line approach to reach acceptance.

**Patterns to Reach Community Acceptance**

When our example procedures are plotted on a 2D chart, where the x-axis is researcher accessibility and the y-axis is difficulty of RCT application, there are clusters in 3 quadrants (Fig. 8). Based on location in the plot, the procedures could be divided into 3 categories, and based on findings of the impact of accessibility and RCT use discussed so far, 3 general archetypes of development can be described.

Type 1 development occurs with investigation of a procedure that is widely accessible to investigators and does not lend itself well to RCT study. Prior to community acceptance of efficacy, many authoring groups publish numerous papers, mostly cohort studies, and a large number of patients are investigated. Type 2 development corresponds to procedures with limited researcher accessibility and unlikely RCT use. Fewer authoring groups study a lower total number of patients with a lower total number of studies published prior to community acceptance, though the majority of investigations are cohort studies. Type 3 development occurs when a procedure has limited accessibility and lends itself to RCT investigation. Like Type 2, fewer authoring groups publish fewer studies with...
fewer total patients, but the majority of the publications are RCTs.

These descriptions are based on the archetype of each quadrant, as if the data point was in the quadrant’s center. In reality, procedures fall at different points along the spectrum and their development may not perfectly correspond to the quadrant’s center, though the procedure does share traits most with the archetype in the same quadrant. Also, there are likely additional influences that could play roles of varying degrees in development of different procedures, resulting in further deviation from the archetypes.

Still, organization into archetypes based on these 2 major factors provides an opportunity to better understand surgical development. One unexpected finding is that Type 2 procedures, although fewer and less accessible, reach PSA more rapidly than more accessible Type 1 procedures (p = 0.04). This is counterintuitive, since it seems logical that a more accessible procedure would be more widely studied and enough data to make a conclusion on the question of efficacy would be collected faster. But it appears that the community accepts less-accessible Type 2 procedures faster with less total data collected, even though the studies are mostly of the same design (cohort studies) as seen with Type 1. There are several potential explanations for this occurrence.

One explanation is that official FDA endorsement speeds community acceptance. Regulatory rulings can have a major impact on publication patterns, as can be seen in the case of VNS for resistant depression. There were 2 major regulatory rulings in the case of VNS for depression. First, the FDA approved the device for use in resistant depression in 2005, then, 2 years later, the CMS issued a National Coverage Determination concluding that the treatment was not effective and CMS would not cover its use. To observe how these decisions impact publication rates on the topic, the annual number of publications on the topic can be seen in Fig. 7A. There is steady growth in the number of publications of both initial and refining studies until FDA approval in 2005 (Fig. 7A first dotted line), indicating growing interest in the topic. After approval, efficacy investigations decline and refining studies maintain a high level for 2 more years until the CMS decision (Fig. 7A second dotted line). After the CMS decision, investigations quickly plummet to the lowest level observed. This supports the view that regulatory decisions impact researcher interest and publication patterns. This precipitous decline in research, of both refining and investigating publications, also serves as an example of how the PSA model changes in response to a procedure being deemed ineffective. It seems likely that the CMS decision itself caused the decrease in publications, not the scientific evidence, since the conclusion was based on evidence that was published years earlier.

On the other hand, for several procedures, PSA was reached and the increase in refining investigations occurred prior to any regulatory rulings. There is a second explanation for why less accessible procedures gain PSA faster with less total data than more accessible procedures, even though the quality of data are the same. There is limited access to the procedure during Type 2 development, and only a small proportion of investigators primarily have access. These researchers publish not just efficacy studies but their own refining investigations before the device is widely accessible. Some of these refining publications are warranted, like a report of improvement in outcomes under certain conditions, but others are premature, such as studying the procedure in a new subpopulation prior to the device demonstrating any broad efficacy. In essence, some investigators appear to embrace the new procedure based on their early observations before multicenter results could be evaluated or the whole field has access to test the procedure. With exclusive access, they can move onto refinements prematurely. This results in an early shift in development, so that by the time the procedure is more accessible, new research groups appear to move directly into refinement, even if efficacy has yet to be objectively demonstrated.

While this occurrence is often benign and does not have any major impact if the procedure’s effectiveness is ultimately verified, it can be damaging if the community embraces a procedure that later proves to be ineffective. This can be seen in the case of VNS for depression. Even though the RCT results did not justify enthusiastic FDA approval, and serious safety concerns regarding the device were raised, the FDA advisory panel vote was 5–2 to approve the device after practitioners and patients offered testimony and pointed toward positive observational study findings. The FDA is not required to accept the recommendation of the expert panel and appeared to be leaning against approval until the director of the FDA’s Center for Devices and Radiological Health overruled objections to approve the device. This is an example in which the relevant community prematurely embraced a procedure, and when it came time to decide its usefulness, expert opinion overruled scientific study.

Managing Scholarly Acceptance

Premature acceptance of an ineffective procedure could result in years of overuse that is harmful to patients and wasteful of resources until the course is corrected. While there is no guarantee that scenarios like this can be prevented, managing PSA to ensure it does not get ahead of objective efficacy could be beneficial. This could be achieved via journals with a robust peer-review process that are more alert to avoiding acceptance of premature refining studies, which may falsely send a signal that a procedure has proven to be efficacious. Likewise, it could be attained by increasing access to researchers, so that skeptics can test early claims before opinions become ingrained.

Community Acceptance May Depend on Clinical Efficacy

Another interesting situation is when the opposite occurs: PSA arrives considerably after objective efficacy. This was the case with Gliadel for the treatment of malignant gliomas. The FDA approved the drug for recurring malignant gliomas in 1996 (and for newly diagnosed malignant gliomas in 2003) [FDA Highlights of Prescribing Information Reference ID 3358686; http://www.accessdata.fda.gov/drugsatfda_docs/label/2013/020637s026lbl.pdf] and (US Food and Drug Administration: Drug Approval Package, Application No: 20-637/S16; http://www.accessdata.fda.gov/) overruled scientific study.
Significance in its current state. Even the Phase III trial conducted in 2003 that led to Gliadel’s approval for newly diagnosed malignant gliomas showed only a modest benefit in median survival (13.9 months compared with 11.6 months for the placebo group), and no significant benefit was found in the patient subgroup that had a histological diagnosis of glioblastoma. So it is possible that a long delay in community acceptance after objective efficacy stemmed from demonstrated statistical significance with doubt of clinical significance in its current state.

Conclusions

Our findings are based on a limited sample size and specific patterns that were identified retrospectively. Future research is required to investigate whether these patterns remain consistent with additional examples. Still, most of the interpretations make sense and our suggested new terminologies can contribute to a better understanding of surgical development. It is important to understand how innovations currently move from conception to acceptance. This will be essential to improving the developmental process for our patients.

References

29. Valtonen S, Timonen U, Toivanen P, Kalimo H, Kivipelto L,


Author contributions
Conception and design: both authors. Acquisition of data: Schnurman. Analysis and interpretation of data: both authors. Drafting the article: both authors. Critically revising the article: both authors. Reviewed submitted version of manuscript: both authors. Approved the final version of the manuscript on behalf of both authors: Kondziolka. Statistical analysis: Schnurman. Administrative/technical/material support: Kondziolka. Study supervision: Kondziolka.

Supplemental Information
Companion Paper

Correspondence
Douglas Kondziolka, Department of Neurosurgery, NYU Langone Medical Center, 530 First Ave., Ste. 8R, New York, NY 10016. email: douglas.kondziolka@nyumc.org.