Off-label innovation: characterization through a case study of rhBMP-2 for spinal fusion

Zane Schnurman, BA, Michael L. Smith, MD, and Douglas Kondziolka, MD

Department of Neurosurgery, NYU Langone Medical Center, New York, New York

OBJECTIVE Off-label therapies are widely used in clinical practice by spinal surgeons. Some patients and practitioners have advocated for increased regulation of their use, and payers have increasingly questioned reimbursement for off-label therapies. In this study, the authors applied a model that quantifies publication data to analyze the developmental process from initial on-label use to off-label innovation, using as an example recombinant human bone morphogenetic protein 2 (rhBMP-2) because of its wide off-label use.

METHODS As a case study of off-label innovation, the developmental patterns of rhBMP-2 from FDA-approved use for anterior lumbar interbody fusion to several of its off-label uses, including posterolateral lumbar fusion, anterior cervical discectomy and fusion, and posterior lumbar interbody fusion/transformaminal lumbar interbody fusion, were evaluated using the “progressive scholarly acceptance” (PSA) model. In this model, PSA is used as an end point indicating acceptance of a therapy or procedure by the relevant scientific community and is reached when the total number of peer-reviewed studies devoted to refinement or improvement of a therapy surpasses the total number assessing initial efficacy. Report characteristics, including the number of patients studied and study design, were assessed in addition to the time and pattern of community acceptance, and results compared with previous developmental study findings. Disclosures and reported conflicts of interest for all articles were reviewed, and these data were also used in the analysis.

RESULTS Publication data indicated that the acceptance of rhBMP-2 off-label therapies occurred more rapidly and with less evidence than previously studied on-label therapies. Additionally, the community appeared to respond more robustly (by rapidly changing publication patterns) to reports of adverse events than to new questions of efficacy.

CONCLUSIONS The development of off-label therapies, including the influence of investigative methods, regulation, and changing perspectives, can be characterized on the basis of publication patterns. The approach and findings in this report could inform future off-label development of therapies and procedures as well as attempts to regulate off-label use.

http://thejns.org/doi/abs/10.3171/2016.1.SPINE151192

KEY WORDS innovation; off-label; development; science
Payers have increasingly questioned paying for off-label therapies, and others have encouraged increased oversight by the FDA or institutional review boards. Surgeons commonly use off-label medical devices and implants, and regulations or restrictions could significantly impact clinical practice and surgical innovation. To better understand the effect of such regulations and inform any discussion of their implementation, it is essential to understand how off-label therapies develop and how the surgical community comes to embrace new off-label uses.

In this study, we analyzed the development of off-label therapies with a model called progressive scholarly acceptance (PSA), which uses publication data and patterns to characterize the process of development from on-label use to off-label innovation. Recombinant bone morphogenetic protein 2 (rhBMP-2), originally approved for use in anterior lumbar interbody fusion, was selected to serve as a case study of off-label use because of its wide utilization for off-label indications in addition to its large clinical impact and high-profile investigations. Between 2002 and 2007, at least 85% of procedures using rhBMP-2 were for off-label applications. Detailing the development of rhBMP-2 and comparing the results to previously studied procedures could help inform our understanding of the process of development, use, and innovation of treatments and procedures.

**Methods**

Bone morphogenetic proteins are growth factors that can induce the formation of new bone and have been used as an alternative to autologous bone graft in spinal fusions. The genetically engineered human protein rhBMP-2 was approved by the FDA in 2002 as a component of the INFUSE bone graft/LT-CAGE lumbar tapered fusion device for the specific indication of anterior lumbar interbody fusion (ALIF). Within the next few years, rhBMP-2 was being applied off label to assist with a variety of fusions, most notably posterolateral lumbar fusions (PLF), posterior or lumbar interbody fusion (PLIF), transfemoral lumbar interbody fusion (TLIF), and anterior cervical discectomy and fusion (ACDF). We used the PSA model to characterize the development of rhBMP-2 for each of these indications.

**Progressive Scholarly Acceptance**

The theory behind PSA and techniques for modeling it are explained in detail elsewhere. In brief, the goal of a PSA model is to use peer-reviewed publications to identify a standard end point when a procedure can be considered to have been accepted by the relevant scientific community.

The idea of PSA is based on the historically supported view that surgical innovators are “tinkerers” who continually refine procedures, and the history of surgical innovation can be broken into a series of broad innovations followed by periods of refinement. Key to the theory of PSA is the recognition that prior to improvement comes acceptance of the broad, initial innovation. That is, the investigators believe in the procedure they are building upon before contributing a refinement.

As a procedure demonstrates efficacy, utilization increases, and more authors begin to study and publish their own refinements of the innovation. At some point, the cumulative number of refining studies eclipses the cumulative number of efficacy studies. This is the point designated as PSA, and it is an indication of community acceptance; the majority of the publishing community has then progressed beyond the initial question of efficacy and moved on to refinement.

It is important to note that the attainment of PSA does not always mean that the efficacy of the innovation has been objectively demonstrated. While these events usually correspond, occasionally the community may believe in the safety and efficacy of the innovation earlier or later than its objective validation.

**Data Collection and Analysis**

To measure PSA for rhBMP-2, a wide net was cast to identify as many studies investigating rhBMP-2 and spinal fusion as possible (Fig. 1). The terms “(bone morphogenetic protein 2) AND Fusion” were searched in PubMed (www.ncbi.nlm.nih.gov/pubmed/) with mapping to medical subject heading (MeSH) vocabulary. Additional filters were applied to limit the review to human studies written in English. All research studies were reviewed in chronological order, from earliest to most recent, through the year 2014. Each relevant report was characterized as either an “initial investigation,” which would be answering the question of broad efficacy, or a “refining study” of one (or multiple) of the procedures being investigated. Examples of refining studies include investigations of modified procedure methods, new indications for the procedure, and uses of the procedure in more narrow populations and in particular disease subtypes. In the case of rhBMP-2 for ALIF, initial studies were early reports from the FDA-approved study investigating rhBMP-2 in the LT-CAGE device, while the earliest refining studies, published by some of the same authors, focused on the application of rhBMP-2 in threaded cortical allografts instead of cages. For more detail on differentiating initial and refining studies, see Schnurman and Kondziolka, 2015.

In addition to publication dates and categorization of the studies as initial or refining, the number of patients studied and the study design were noted to further characterize developmental patterns. The values for total number of patients studied (controls not included) can only be considered as estimates of total patients studied, as studies may have included follow-up data or more specific investigations of the same patient population in later studies (resulting in double counting). The mean 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 randomized controlled trial (RCT). The classification of observational studies was based on the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist, with additional specificity regarding the differentiation of case-series and cohort studies based on Dekkers et al., 2012, with cohort studies enabling absolute risk and outcome predictions based on clinical data.

For all data, the starting point was the first modern J Neurosurg Spine Volume 25 • September 2016 407
publication on the topic and the end point was the year when PSA was reached. Characteristics were analyzed for both initial and refining studies, although emphasis was placed on outcomes from initial studies because these are the investigations that test initial efficacy and should indicate the burden of evidence required for community acceptance.

Assessing Conflicts of Interest

Disclosures of conflicts of interest (COIs) for all studies from 2000 to 2014 were reviewed. There is no universal standard on reporting COIs across journals, and policies on COI disclosure have evolved over the 14 years investigated in this study, adding difficulty to defining bias. Articles with potential bias include investigations that were fully funded by Medtronic, owner of the rights to INFUSE; studies that had authors receiving benefits related to the studied topic; and articles that had no information on COIs. Rather than delineate different sources of bias, we tabulated articles that explicitly reported no COIs. The difference in the percentage without COIs was compared between initial studies and refining studies using the unpaired t-test.

Progressive Scholarly Acceptance Analysis

The first step in our analysis was to identify the point of PSA, the end point used for community acceptance. Two PSA models were used, an annual model and a cumulative model (referred to as the compounding model in our prior report). In the annual analysis, the number of initial investigations and number of refining studies each year were plotted against time as line graphs. With this model, the annual number of studies can be seen for each year, making it a valuable tool to visualize fluctuations in the number of publications and the immediate impact of different factors. For the end point of community acceptance, however, the more conservative cumulative model was used. It is similar to the annual model—initial and refining studies are plotted against time—except the total number of initial and refining publications are cumulative each year. Thus, if the refining line surpasses the initial line, then the total number of reports on improving a procedure is larger than the total number of reports on studies of its efficacy. For this to occur, publications focusing on continued refinement must outpace those focusing on efficacy. Reaching this point implies the majority of the community publishing on the topic has embraced the procedure, and the priority is procedural improvement. This is the point we call progressive scholarly acceptance, and it is the end point for our analyses.

Results

rhBMP-2 for Anterior Lumbar Interbody Fusion—On-Label Use

The first peer-reviewed publication on the use of
rhBMP-2 in humans was a preliminary report of 11 patients by Boden et al. in 2000, who studied the application of this drug in ALIF in an FDA-approved device exemption RCT. The FDA approved Medtronic’s InFuse bone graft for specific use in ALIF with LT-CAGE for degenerative disc disease on July 2, 2002, based on the full results of a 2-year trial published by Burkus et al. in the same year. (Of note, in addition to ALIF, FDA-approved indications for rhBMP-2 include treating open tibial shaft fractures stabilized with an intramedullary nail and sinus or with localized alveolar ridge augmentations, approved in 2004 and 2007, respectively.

In 2002, the year of FDA approval for ALIF, Burkus et al. and Boden et al. each published studies of the earliest refinements: rhBMP-2 in ALIF using threaded cortical allografts instead of cages and its application in posterolateral fusion.

Over the next few years, new investigators began to study and publish on rhBMP-2 for spinal fusion, most moving directly into refinement. In 2004, the total number of refining studies eclipsed the total number of initial studies (Fig. 2), and 2004 is considered the point of PSA. It took 4 years to reach PSA with 6 initial clinical studies (Table 1). Of these 6 reports, 5 (83%) were RCTs and 1 was a cohort study. These initial reports involved a total of 801 patients, making a mean of 134 patients per study; although it should be noted that many of these reports are based on the same RCT (which had a total of 143 investigational patients).

Data collection continued through 2014. Of note, a decrease in refinement was observed in 2012, likely due to increasing scrutiny of the failure to accurately report complication rates for the original industry-sponsored trial in publications that included the entire June 2011 issue of The Spine Journal dedicated to the topic. There was also a small resurgence in initial investigations assessing basic safety and efficacy, including the assessment by the Yale Open Data Access (YODA) Project in 2013. The impact of these investigations on rhBMP-2 development can be seen in all indications (Figs. 2–5 left in all).

### rhBMP-2 for Posterolateral Fusion—Off-Label Use

The first clinical investigation of rhBMP-2 in the posterolateral spine was a small RCT involving 20 patients published in 2002. These were early results from a trial with FDA support for higher-dose rhBMP-2 and an hydroxyapatite/tricalcium phosphate-collagen carrier (AMPLIFY rhBMP-2 Matrix) in PLF. Multiple updates on the trial were published in 2005 and 2006.

Early refinement had begun at the same time, such as in studies of rhBMP-2 for PLF specifically in adult spinal deformities and the effect of smoking on fusion. Refinement increased and the total number of refinement studies surpassed the total number of initial studies, reaching PSA in 2007 (Fig. 3). This was prior to the publication of the full AMPLIFY trial results in 2009, and the FDA ultimately sent a nonapproval letter for Medtronic’s AMPLIFY bone graft in 2011 because of increased cancer risk from its higher rhBMP-2 dose.

PSA was reached 5 years after the first publication (Table 1). Three initial studies were published prior to PSA, with data from 111 patients (a mean of 37 patients per study). All 3 reports were based on an RCT.

### rhBMP-2 for Anterior Cervical Discectomy and Fusion—Off-Label Use

The first application of rhBMP-2 for anterior cervical fusion was in 2003 by Baskin et al. Within 2 years, multiple refinements were published, such as a trial combining rhBMP-2 with bioresorbable interbody spacers. The total number of refinement studies eclipsed initial studies for a brief period of time (Fig. 4), but more initial studies were published, likely spurred by increasing reports of cervical swelling with rhBMP-2 use, which was first assessed in the literature in 2006. In July 2008, the FDA issued a public health notification warning practitioners about increased risks of using rhBMP-2 in the cervical spine. Despite this warning, refining investigations continued, and the total number of studies refining rhBMP-2 use for cervical spine permanently eclipsed the number of studies assessing initial safety and efficacy in 2009 (Fig. 4).

It took 6 years to reach PSA (Table 1), with the publication of 7 initial studies. Of these 7 studies, 5 (71%) were cohort studies, with 1 RCT and 1 case study. A total of 355 patients were assessed in these studies, for a mean of 51 patients per study.

---

**FIG. 2.** Graphs showing PSA of rhBMP-2 for ALIF, with the annual publication of initial and refining reports (left) and the cumulative number of reports each year (right).
rhBMP-2 for Posterior Lumbar Interbody Fusion and Transforaminal Lumbar Interbody Fusion—Off-Label Use

The first clinical studies of rhBMP-2 applied to PLIF/TLIF were published in 2004, including an RCT of 34 patients comparing the procedures to autologous iliac crest bone graft (ICBG). Over the next 7 years the number of initial and refining studies continued at a steady pace; some years there were more initial studies, others more refinement, until by 2011 the total number of refining studies had surpassed the total number of initial studies (Fig. 5). This is the point of PSA.

It took 7 years to reach PSA (Table 1). There were 8 initial studies, 6 (75%) were cohort studies, 1 was an RCT, and 1 was a case series. A total of 349 patients were studied in these investigations, for a mean of 44 patients per initial study.

Reported COIs
For each procedure, the studies that explicitly reported no COIs were assessed (Table 2). Of the 77 ALIF studies reviewed, 29 articles (38%) reported no COIs, including 35% of initial studies and 39% of refining studies. For PLF investigations, 33% of the 40 studies reviewed reported no COIs, including 23% of initial studies and 37% of refining studies. Of the 28 ACDF investigations, 14 (50%) reported no COI; 36% of initial studies and 59% of refining studies for ACDF reported no COIs. Finally, 45% of the 40 articles reviewed for PLIF/TLIF reported no conflicts, including 50% of the initial studies and 43% of the refining studies. There was no significant difference in the percentage without COIs between initial and refining studies (p = 0.30).

Discussion
The purpose of this investigation was to characterize the development of rhBMP-2 for its FDA-approved use, ALIF, and several of its off-label uses, PLF, ACDF, and PLIF/TLIF. By comparing the findings to previously characterized surgical procedures, the development of rhBMP-2 may be used as a case study for off-label development. This may improve understanding of the patterns of community acceptance for off-label therapies and how they differ from conventional therapies.

Acceptance of Off-Label Therapies
In a previous investigation, PSA was used to characterize high-impact neurosurgical procedures and potential influences on development. Of those procedures, the mean number of initial studies prior to PSA was 13.7 (range 4–25) and the mean number of years to PSA was 9.1 (range 5–13). Neither characteristic followed a normal distribution, but instead clustered into multiple groups depending on 2 key factors: use of RCTs and FDA approval.

When a procedure received FDA approval, it tended to become accepted by the community more rapidly, even with fewer, and not necessarily higher quality, total studies. This makes sense because FDA approval serves as a trustworthy endorsement of efficacy, and investigators may accept the procedure and move on to refinement. Similarly, studying a procedure with RCTs resulted in fewer total studies prior to acceptance. This is also sensible, as higher-quality evidence may lead to more confident belief in efficacy and acceptance, even if fewer total patients were investigated.
Given that procedures utilizing medical devices off-label have not been approved by the FDA and are less likely to be studied extensively with RCTs, it might seem likely that they would follow a slower developmental pattern and require a larger number of published investigations prior to community acceptance, similar to patterns seen with nondevice procedures. But this is not what was seen with rhBMP-2 off-label use.

The on-label use of rhBMP-2 for ALIF reached PSA in 4 years, with 6 initial articles prior to acceptance, mostly reports of RCTs (83%). Compared with the prior study of development, this is on the more rapid spectrum and lower range of articles, but is expected and agrees with previous findings given FDA approval 2 years after the first publication and significant RCT utilization.

The 3 off-label therapies also demonstrated rapid community acceptance with fewer articles, with less clear explanations why. rhBMP-2 use for PLF reached PSA in 5 years with only 3 initial studies. All 3 studies were RCTs, which could explain the rapid acceptance of this therapy, although all were based on a single trial. ACDF was accepted in 6 years with 7 studies that were majority cohort (71%) investigations. Similarly, the use of rhBMP-2 in PLIF/TLIF reached PSA in 7 years with 8 studies, also majority cohort (75%). All of these procedures reached PSA relatively rapidly with few total initial studies.

One explanation is that the community more readily accepts a previously endorsed procedure applied to a new indication than a de novo procedure. When a device or implant has been previously demonstrated as safe and efficacious in stringent RCTs and approved by the FDA, and then is applied to a related procedure, it may be assumed to have similar efficacy and safety. After limited investigation supporting that notion, many clinicians accept the new application and even publish refinements, which may further signal to other clinicians to accept the application. While this makes intuitive sense, and can often be reasonable, it does have significant implications.

A new application of an established therapy may have different risks, as was true with rhBMP-2 when applied to the cervical spine. In July 2008, the FDA issued a public health notification, warning practitioners about increasingly seen instances of neck swelling and potentially lethal airway obstruction with rhBMP-2 in the cervical spine. Community acceptance and utilization of a procedure prior to objectively demonstrated safety and efficacy may be an inherent risk of off-label utilization.

**Influence of Bias**

Another possible explanation for the rapid acceptance of off-label therapies is the behind-the-scenes promotion...
by drug and device companies to educate prescribers of the treatment’s beneficial effect, even without objective evidence. While the FDA prohibits direct promotion of unapproved use, drug companies can indirectly promote use by sponsoring continuing medical education programs and distributing journal articles about the off-label use.20

Provided articles are often industry funded, and investigators may have a financial COI.19 This is of particular concern with rhBMP-2, as the influence of the pharmaceutical company and the bias of investigators have been claimed as the reason that complications from the original trial were not initially reported.9

The disclosure information from rhBMP-2 articles from 2000 to 2014 was assessed. For investigations of ALIF, PLF, ACDF, and PLIF/TLIF, 38%, 33%, 50%, and 45%, respectively, explicitly reported no COIs. From this limited sample, there were no clear differences in level of bias between initial and refining studies. Additional studies are necessary to further characterize the impact of external funding and investigator bias in treatment development.

The Effect of Technology-Related Morbidity

Too rapid acceptance of a procedure, which may be more likely for off-label therapies, may result in overlooking adverse events and having to respond to them in a delayed fashion. After a procedure has reached PSA and is widely used, there appear to be 2 main reasons for a reversal in views and a movement back toward efficacy studies: 1) new questions of efficacy and 2) newly demonstrated complication profiles. The community’s responses to these 2 issues appear to be different.

Newly questioning whether a procedure is efficacious, after it has already been accepted, can lead to a slow reversal of opinions without a more significant influence, such as a regulatory ruling. This was seen in a previous study of vagus nerve stimulation for treatment-resistant depression.27 After the largest RCT to date reported limited efficacy, the community appeared to disregard the result and continued refinement until the Center for Medicare and Medicaid Services issued a National Coverage Decision rejecting payment 2 years later, which resulted in refinement cessation.27

On the other hand, the patterns seen with rhBMP-2 indicate that the response to reports of morbidity may be more robust. After early reports of adverse cervical swelling with rhBMP-2 use in 2006, a rapid reversal between initial study and refinement in ACDF can be seen as authors moved to further investigate safety (Fig. 4 left). This is reasonable since harm is more concerning than lack of benefit. When the morbidity risk is raised, investigating those risks becomes a priority, and publication of initial studies reviewing safety and efficacy overtakes refinement.

The decrease in refinement research may be only temporary if the complication can be better characterized, as was seen with rhBMP-2 for ACDF. Two years later, initial studies decreased as refinement increased again in 2008 (Fig. 4 left). This is likely because several years of further initial study had better described rhBMP-2–induced cervical swelling, and authors attributed the risk mostly to dosage and delivery.17 After better defining causative factors and preventative methods, authors may feel more comfortable managing adverse event risks and return to refinement.

A similar robust decrease in refinement can be seen with rhBMP-2 for ALIF. Controversy erupted in response to questions of bias and underreporting of adverse events, such as retrograde ejaculation, in the original industry-sponsored investigations.9,10 A rapid decrease in refinement can be seen in 2010–2012, the peak time of controversy surrounding rhBMP-2 (Fig. 2 left).

Spurred by the controversy, the YODA Project was commissioned to provide the original rhBMP-2 clinical trial data to 2 independent research teams for review. In 2013, 2 meta-analyses were published, reporting that rhBMP-2 and ICBG had no clinical difference in fusion rates and had similar complication rates, including retrograde ejaculation.15,23,29 After better understanding the true risks of these complications, it would be expected that refining would again increase, as was seen in the investigation of rhBMP-2 for the cervical spine. Indeed, there appears to be a resurgence of refinement in investigation of rhBMP-2 for ALIF, starting in 2013 (Fig. 2 left), though it may be too early to say conclusively whether this trend will continue given that data were only collected through 2014.

Conclusions

There are significant benefits of off-label utilization, such as increasing patient access to innovative technologies; however, more readily accepting these therapies with inadequate evidence comes with inherent risk, and responding to adverse events after widespread clinical utilization is not ideal. To maintain benefits and manage risk, off-label use should be accompanied by thorough reporting of outcomes through a variety of forms, including patient registries, RCTs, and observational studies.6,23 Journals, via a robust peer-reviewed process, can alert authors and readers to avoid premature refining studies, which can help to ensure that procedure PSA does not get ahead of objective efficacy. Additionally, a key to ensuring safe off-label use will be better understanding and managing the influence of pharmaceutical funding and the impact of bias on acceptance of new treatments. Understanding the patterns of off-label development is essential to informing regulations and improving the developmental process.

References


17. Glassman SD, Dimar JR, Carreon LY, Campbell MJ, Puno RM, Johnson JR: Initial fusion rates with recombinant hu-

Disclosures
Dr. Kondziolka reports receiving research funding unrelated to this study from Brainlab, Inc., and Elekta AB.

Author Contributions
Conception and design: all authors. Acquisition of data: Schnurman. Analysis and interpretation of data: all authors. Drafting the article: all authors. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Kondziolka.

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