Introduction of New Devices and Technologies Into a Spine Surgery Practice: A Review of Processes and Regulations

Sigurd H. Berven, MD; Michael J. Yasemski, MD, PhD; Peter O. Newton, MD; William Christianson; Harold M. Aberman; Jean-Charles Moreau; Mary Jane Mulcahey, PhD; Randal R. Betz, MD

Informed decision making in the development and use of new technologies requires a thorough understanding of the data that support their approval. With education, physicians are empowered to adopt clinically directed applications rather than be restrained by strict on-label usage.

The emergence of new devices and new technologies has led to significant and important changes in the practice of spine surgery over the past decade. Spine surgeons are responsible for continuously learning about new implants, instruments, bone graft substitutes, and new procedures. Making informed decisions on the adoption of new technologies in spine surgery is an important challenge for all practicing surgeons. This article provides a systematic review of the processes involved in bringing new technologies and devices to market in spine surgery. It is intended to assist surgeons in building a knowledge base to evaluate whether the new options are appropriate for their patients and/or spine practice and to critically evaluate the processes of evidence development to facilitate informed decision making about new technologies.

REGULATORY APPROVAL FOR NEW DEVICES AND TECHNOLOGIES

Safety has been addressed by the US Food and Drug Administration (FDA) through classification of medical devices as Class I, II, or III. The failure of pacemakers and intrauterine devices and the lack of medical device regulation in the early 1970s led to the Medical Device Amendments of 1976. Class I devices (eg, tongue depressors) fall under general controls that apply to all devices, such as registration of manufacturers, record keeping requirements, and compliance with good manufacturing regulations. Class II devices (eg, most orthopedic implants) are those for which general controls are insufficient to ensure device safety and effectiveness, and for which ≥1 special controls may be required. These special controls may include performance standards, postmarket surveillance studies, or device registries. Class III devices (eg, pacemakers, heart valves) are those that are life supporting, life sustaining, or important in preventing impairment of human health. General and special controls may be insufficient to ensure the safety and effectiveness of a Class III device, and it would be required to go through premarket approval via an investigational device exemption clinical study.

Generally, a Class I device carries minimal risk (eg, non-powered surgical instruments). A Class II device presents medium or manageable risk and
usually has a predicate device (one that was marketed prior to the May 1976 medical device amendments and has an extensive history). Pedicle screws are an example of a Class II device in that they only need biomechanical testing and to follow material standards to show that they meet basic mechanical properties, without further proof of efficacy before marketing. Class III devices carry the potential for high risk or have no antecedent technology to permit a comparison. Class III devices require both safety and efficacy data before market release.

The classification of new devices is important in determining the process of bringing a new device to the marketplace. A new technology that is substantially equivalent to existing devices that have been in use since prior to 1976 may be cleared for marketing by the FDA under the process of 510(k) clearance. The FDA 510(k) regulatory process specifically applies to devices and technologies for which there was a predicate device in existence prior to the May 1976 medical device amendment or for a device found to be substantially equivalent post-1976 that has been cleared. For 510(k) approval, the FDA typically requires a device description, citation of a predicate device, mechanical drawings, bench testing, and a package insert. Therefore, many new technologies may qualify for FDA approval without demonstration of efficacy compared with alternatives and without specific clinical trials. When the heritage of a new device is substantially equivalent to a device that existed prior to 1976, or the device has an established safety and efficacy history, it is generally classified as Class I or II and may bypass clinical trials before approval for marketing.

Some products may be approved by a 510(k) process and then be used in a physician-directed, or off-label, application. Physician-directed application of a device or medical product is a clinical use that is not specific to the product approval. The FDA does not regulate medical practice. Some surgeons believe that the term “off-label” is an inappropriate term to refer to a decision that is part of medical practice, since labeling is an activity that occurs between the FDA and industry. We suggest adoption of the alternate term “clinician- or physician-directed application” as a more appropriate description for this activity.

Class III devices and technologies, or devices without a substantially equivalent precedent, require a rigorous process of evaluation for which the FDA will require preclinical studies and clinical trials as well as manufacturing and biomechanical data. The clinical trials may include pilot studies and a level of evidence such as a randomized, controlled trial. Examples of Class III spinal devices include artificial disk replacement devices, prosthetic nuclear replacements, and recombinant bone morphogenetic proteins (BMPs). The premarket approval process is rigorous and results in a level of evidence that provides the practitioner useful data on efficacy as well as safety.

The premarket approval process begins with a meeting between the sponsor of the premmarket approval application (industry/manufacturer) and representatives of the Office of Device Evaluation. The purpose of this meeting is to design preclinical and clinical studies that will demonstrate product safety and ensure efficacy. Clinical data are obtained from investigational device exemption studies, which may involve multiple centers, to provide generalizability of product application and extensive datasets. The determination of primary and secondary outcome measures and definitions of study success are critical to study design and to eventual assessment of product performance. Food and Drug Administration reviewers evaluate the investigational device exemption study and determine the merit of the application. A Product Development Protocol is an alternative to the premarket approval. The Product Development Protocol pathway includes preclinical and clinical evaluations and may reduce the burden of the approval process by involving the FDA in the earliest stages of study planning. After approval of a new device by the FDA, postmarket analysis, including reporting adverse events and annual reports, may be required for ongoing quality assurance.

Burden of Proof and Physician Adoption of a New Technology

Efficacy is an important consideration for physicians in choosing to adopt a new technology. Efficacy may be supported by preclinical data. The most useful preclinical data on efficacy is a matched trial of similar products in an animal model. Information on new bone graft substitutes is often limited to preclinical data, and surgeons should not make decisions based solely on preclinical data. Orthotropic sites and phylogenetics of the animal model are important in determining the level of proof in preclinical work. Clinical data, including prospective, randomized trials, are more useful for a determination of efficacy. However, this level of proof is generally only available for new products subject to the premarket approval process and investigational device exemption trials (Class III devices).

While Class III devices are required to demonstrate efficacy prior to approval, physician-directed (off-label) applications may also limit the quality of evidence that supports clinical use. An example in spine surgery is the use of recombinant BMPs. Bone morphogenetic proteins have been shown to be safe and effective in anterior interbody devices in a 1-level anterior spinal fusion (BMP-2). Many surgeons adopted the use of BMP-2 in the posterolateral spine despite the fact that the preclinical data did not support its use without the appropriate carrier or in the dose used in the anterior interspace. Barnes et al show in a primate model that BMP-2 with a compressible matrix is not effective as a bone graft substitute in the posterolateral space. Despite this preclinical data, BMP-2 is used most commonly in posterolateral applications based on limited clinical data. Through ongoing postmarket analysis, the efficacy of BMP-2 with a noncompressible matrix in the posterolateral space has been demonstrated.
This example points to the importance of ongoing postmarket analysis in industry-sponsored and independent investigations. It is incumbent on the surgeon to understand the level of evidence that may support or refute specific clinical applications of new technologies and to continue to study efficacy of specific applications in postmarket studies. Numerous examples exist where initial clinical trials may suggest efficacy but follow-up investigations in independent studies fail to demonstrate a significant effect. Examples in spine surgery include chemonucleolysis, injections and nerve blocks, and intradiscal electrothermal therapy. Ongoing clinical trials even after FDA approval of devices are important to demonstrate efficacy and to guide informed choices regarding care.

Cost is the third element that may contribute to a physician’s choice to make a change in practice. An increase in the cost of a new technology that also results in improved outcomes most directly affects the patient and the physician. The benefit to the hospital and the payer is not as clear. The most useful measure of the value of a new technology is the incremental cost effectiveness ratio, or the ratio of cost per unit of outcome (use) gain. The threshold of incremental cost effectiveness ratio that justifies the adoption of new technologies for all stakeholders has not been well established.

**Clinical Study Design**

Preclinical data may serve to compel a change in practice, or more often may provide data that may support a change in practice. Clinical studies are the currency of assessment of efficacy for new devices and technologies, and the design and implementation of these studies is fundamental to approval and reimbursement for new technologies.

A unique problem for surgeons, companies, and the FDA is the cost of an investigational device exemption premarket approval trial, which can be prohibitive. An excellent example of this is the growing rod. Whereas surgeons are using this as treatment for progressive curves in early onset scoliosis, there is no level I evidence to support its efficacy, nor is there explicit FDA approval for this indication. The surgeons themselves could conduct a quality-controlled study in the form of a surgeon-driven investigational device exemption, but the prohibitive costs render this option impractical. Surgeons often make practice decisions based on incomplete data and lower-level studies including case series, retrospective cohort studies, and peer opinion and recommendations. Generally, the surgeons can analyze retrospective cohorts for comparison, but these studies can be fraught with missing data, lost patients, selection bias, and limitations that are not present in sponsored, prospective investigational device exemption trials.

A similar dilemma involves other fusionless technologies, specifically the staple for the young patient (aged 8 to 14 years) with idiopathic scoliosis. When considering the cost of a company doing an investigational device exemption premarket approval trial comparing stapling to a brace, the power analysis shows that the numbers necessary to prove superiority are so large that a study would not be financially feasible. In 1996, the FDA instituted a new classification called humanitarian device exemption, which applies to devices for which the projected use is <4000 cases per year. Devices may be approved with an humanitarian device exemption showing safety but without the level of efficacy data that would be required by a full premarket approval because the condition treated, or the application of the product, would not be large enough to support a randomized clinical trial. An example of an humanitarian device exemption device is the Vertical Expandable Prosthetic Titanium Rib (Synthes Spine, Paoli, Pennsylvania). No precedent technology exists that can serve as an adequate control group of young patients with complex spine deformities. Observation or fusion in situ is not acceptable in many cases of early-onset scoliosis. The humanitarian device exemption process enables surgeons and industry to collect safety and clinical efficacy data for these rare applications of new technology.

The limitation of humanitarian device exemption approval is that surgeons are required to obtain Institutional Review Board approval to use the device at their institutions. The company producing the device cannot make a profit on the device and may only charge enough to recoup the cost of manufacturing and distribution, development, education of surgeons to use the product, and further research on the outcomes of product use, thus limiting the financial attractiveness of this regulatory pathway. One further important restriction is that surgeons can only use an humanitarian device exemption-approved device in a physician-directed application if the situation in which it is needed is immediately life threatening and no other treatment exists, and this scenario would likely never occur in spine surgery.

Alternatives to randomized prospective clinical trials may be less costly and more practical for devices that have limited and highly specialized use. Rudicel and Esdaile support a surgeon randomization process in which the enrollment (number of patients) can be controlled in each. If the inclusion criteria are the same, this theoretically eliminates the bias of the surgeon choosing 1 procedure over the other and preselecting what he or she thinks might be best for a given patient. If the surgeon does just 1 procedure but different surgeons do different procedures, this may be scientifically valid. While still not as good as a randomized controlled trial, it eliminates some bias. One potential bias to the study, however, is asymmetry in surgeon performance. This effect may be significant for devices that have a steep learning curve.

The industry perspective on investigational device exemption studies and new product introduction to the marketplace is different than the perspective
of the surgeon or the payer. The companies may use the resulting information to market a new device/technology or compare it to existing devices. Conducting a formal clinical investigational device exemption premarket approval is financially risky for a company. Therefore, the company may seek approval via the 510(k) process rather than the premarket approval process if there exists a comparable precedent device.16 One of the interim steps a company may take is to apply for a European Conformity (CE) mark to allow marketing a product in Europe.17 The CE mark is considerably easier to obtain than a premarket approval because of the distinction between Class IIb devices in Europe and Class III devices in the United States. For example, in Europe, interbody cages are classified as IIb and, therefore, are approved with reliance on clinical data from other interbody fusion cages. As a consequence, in Europe there is minimal physician-directed (off-label) use of devices as compared to the United States.

Whereas a CE mark may take 6 months to obtain, an investigational device exemption premarket approval may take several years. Even if the company has plans to do an investigational device exemption premarket approval, the data from outside the United States can help the company in the US approval process in several ways, including providing information for a statistician to develop a power analysis that may limit the number of patients needing to be enrolled in an investigational device exemption study.

In addition, if the outside-US data are collected and quality controlled in a superior fashion, it may be acceptable to the FDA for approval. Several companies are collecting good quality safety and efficacy data outside of the United States because of the cost being considerably less than a US clinical investigational device exemption trial.

The endpoint of a premarket approval study is to demonstrate safety and efficacy relative to metrics that are based on an existing treatment and agreed to by the sponsor and the FDA prior to beginning the investigational device exemption premarket approval clinical study. While some trials are designed to demonstrate superiority, this endpoint may not be a reasonable goal for products with smaller sizes or incremental efficacy. Noninferiority trials are common because of the power analysis and the ability to financially control costs in a study. The advantage of a noninferiority trial is that the company may obtain regulatory approval from the FDA or other regulatory body with a smaller study design. However, the noninferiority outcome may not provide strong enough efficacy data to support reimbursement by a third-party payer. This was the unfortunate experience of the early premarket approvals for lumbar disk arthroplasties. The role of the third-party payer is important in the eventual marketing and implementation of new technologies.

**The Role of Reimbursement**

Third-party payers are important stakeholders in the adoption of new technologies in spine surgery.18 After a product is approved by the FDA, demonstrating safety and some degree of efficacy, third-party payers, including the Centers for Medicare and Medicaid Services and private insurers, make decisions on whether to reimburse a specific procedure or device and how much to reimburse relative to precedents.19,20 The third-party payer views reimbursement decisions from the perspective of overall cost of care, generally in the short to intermediate term. Incremental cost-effectiveness studies and clinical outcomes expressed in units of health state may align the perspective of the payer with that of the patient and physician more effectively than analysis of cost alone. The Centers for Medicare and Medicaid Services is prevented from using cost-effectiveness data explicitly in making coverage decisions, but private payers may take cost-effectiveness data into consideration in coverage and pricing decisions.

Reimbursement decisions may have a significant impact on physician practice.21 In scenarios of clinical equipoise, reimbursement considerations may drive surgeon behavior in the absence of clear evidence between alternatives. Examples of reimbursements that may influence behavior include a separate current procedural terminology code for prosthetic interbody cages compared with structural allograft, the use of interbody fixation in lumbar arthrodesis, and the use of a plate in single-level anterior cervical disectomy and fusion procedures. An alignment of reimbursement strategies with clinical outcomes and an understanding of incremental cost effectiveness may be the most effective way to promote evidence-based care.22

In the United States and Europe, payers are demanding a higher level of efficacy than that required by the regulatory approval bodies. Whereas one CE mark covers all 29 countries, each country has its own reimbursement approval process. In the United States, outside of the Centers for Medicare and Medicaid Services, insurance carriers and independent organizations play an important role in the assessment of new technologies and in the decision regarding approval and reimbursement.23 Whereas there is a perception that the reimbursement bodies are anti-innovation, they should be perceived more as providence.

In the United Kingdom, the private insurers require more evidence of benefit than the national health insurance requires. One of the options for surgeons is to develop long-term registries. These registries should be designed to document validated outcome measures, including quality of life, and not just administrative measures that will be used by the reimbursement agencies, such as length of stay and short-term process measures.24 This development of long-term registries is going to be a challenge for all surgeons and should ideally be promoted by spine societies. The British Scoliosis Society was asked about compliance of data entry by surgeons within their society, and it is considered to be extremely poor. In the United
The responsibility of the surgeon to determine the required level of proof for a new technology, and useful comparisons of products may be made in comparing the highest level of proof available for similar products. Preclinical data that is of most value include studies done in orthotopic sites (spine-specific) on animals in higher phylogenetic categories. Physicians need to know that there is limited translational validity from rodents to larger mammals and humans.

GLOBAL CONSIDERATIONS IN NEW PRODUCT REGULATION

Requirements for device approval are significantly different between the United States and Europe, in both pre- and postmarket phases. A Class III classification is different in the United States than it is in Europe. In the United States, both an implanted defibrillator and a new type of spinal implant are considered Class III if a device was manufactured after 1976 and has no predicate. In Europe, depending on the applications, the long-term implantable devices can be classified as Class IIa, Class IIb, or Class III. For example, artificial disks are classified IIb in Europe and not Class III as they are in the United States. Therefore, new technologies including the artificial disk may be adopted in Europe with preclinical data that may be equivalent to a 510(k) process rather than a full investigational device exemption clinical study.

The CE mark system (The European Directive on Medical Devices #93/42), which has been in force since 1995, defines provisions with regard to the safety and performance characteristics of medical devices and the health protection of patients and users. Following the duration of use, risk, and indications, the directive defines device classification. Devices have been classified from Class I (low risk, like reusable surgical instruments) to Class III (high risk, like biomaterials, heart valves, and stents). Most spinal implants are in Class IIb.

While a clinical study is not necessary for a product to be approved as Class I or Class II, the relevant scientific literature must be compiled and critically evaluated. For a Class III device, safety and performance are generally demonstrated with a clinical study. In addition, the design dossier is reviewed by the Notified Body before the CE mark is granted. Devices other than custom-made or those intended for clinical investigations are required to meet the essential requirements of the Medical Devices Directive and must bear the CE mark of conformity when they are placed on the market. The manufacturer bears full responsibility for products placed on the market. In line with this new approach, CE marking is a passport for medical device circulation within all the European Union countries.

Since May 1, 2004, the European Union has regulated the status of spinal implants within the 25 member states, which represent >450 million people. Europe has a larger population than the United States, and each country has its own culture and language. The European Union continues to grow, with 2 new member states (Bulgaria and Romania) joining on January 1, 2007. European regulation is also followed by members of the European Economic Area that includes Switzerland, Norway, Iceland, and Liechtenstein. Today, the European Medical Devices Directive is followed by 31 countries.

The European Vigilance Program is a postmarket surveillance obligation in Europe with a medical device vigilance system. This vigilance system was set up to minimize risks to patients by reducing the likelihood of serious incidents involving a medical device. This is achieved through the evaluation of reported incidents by the member states, through the dissemination of information that could be used to prevent a recurrence of the incident or to alleviate the consequences of such incidents and, where appropriate, by the device being modified or taken off the market.

The manufacturer or its representative is legally bound to report to authorities (1) any serious or potentially serious incidents involving devices they produce or sell and (2) systematic recalls of the device to prevent risk of serious injury or death presented by its use. The definition of a serious incident is “any malfunction, failure, or deterioration in the characteristics or performance of the device, or any inaccuracy in the labeling of instructions for use that might lead to or might have led to the death of the patient or user or to a serious deterioration in his state of health.” The CE mark system, with the subclasses IIa and IIb, has features that may be more prag-
matic than the FDA's approval system. For example, in the United States, prior to 2006, interbody cages labeled for use as a vertebral body replacement were in Class II and subject to approval via the 510(k) process. However, the majority are used in physician-directed applications in the interbody space. Until recently, cages labeled for use in the interbody space were in Class III and required premarket approval for marketing. Four US companies have undertaken the expense of conducting investigational device exemption studies and obtaining premarket approval, but the majority have chosen the 510(k) route. In Europe, interbody cages are classified as IIb and therefore are approved with reliance on clinical data from other interbody fusion cages. As a consequence, in Europe there is minimal physician-directed (off-label) use of devices as compared to the United States, where there is extensive use of devices for indications that are different than those approved by the FDA.

**Conclusion**

Informed decision making in the development and use of new technologies requires a thorough understanding of the data that support the approval of a new technology or device. With education and information, physicians are empowered to adopt clinically directed applications rather than be restrained by strict on-label usage. Lessons from Europe include the classification of new devices with a reasonable safety profile as Class IIb, enabling a more efficient process of adoption while maintaining rigorous assessment of safety and efficacy. Many spine surgeons feel that the term “off-label use” should be abandoned because it has a negative connotation. They recommend “clinician- or physician-directed applications.” Physically directed applications require a thorough understanding of the safety profile of devices, their preclinical data, and existing clinical study data for their responsible use.

The reimbursement approval process has important implications for the adoption and dissemination of new technologies. Accountability for the outcomes of care extends long beyond the approval of new technologies by the FDA and by reimbursement authorities. A long-term registry recording outcomes measures needs to be developed in a partnership between surgeons, professional societies, and industry to assess the safety and efficacy of new devices and technologies over time.

**References**


