

SUMMARY MINUTES

**MEETING OF THE ORTHOPAEDICS AND REHABILITATION DEVICES
ADVISORY PANEL**

OPEN SESSION

NOVEMBER 21, 2002

**Gaithersburg Holiday Inn
Gaithersburg, MD**

Meeting of the Orthopaedics and Rehabilitation Devices Advisory Panel

Attendees

November 21, 2002

Chairperson

Michael Yaszemski, M.D., Ph.D.

Executive Secretary

Hany Demian, M.S.

Members

Maureen A. Finnegan, M.D.

John S. Kirkpatrick, M.D.

Kinley Larntz, Ph.D.

Sanjiv H. Naidu, M.D., Ph.D.

Consultants

Albert Aboulafia, M.D.

John Doull, Ph.D.

Hari Reddi, Ph.D.

Andrew Schmidt, M.D.

Consumer Representative

Karen Rue

Industry Representative

Sally Maher, Esq.

FDA Representative

Celia Witten, M.D., Ph.D.

CALL TO ORDER

Executive Secretary Hany Demian called the meeting to order at 9:32 a.m. He noted that Albert Aboulafia, M.D., John Doull, Ph.D., and Andrew Schmidt, M.D., had been appointed to temporary voting status. Mr. Demian then read the conflict of interest statement. Albert Aboulafia, M.D., John Kirkpatrick, M.D., Kinley Larntz, Ph.D., and Andrew Schmidt, M.D., had been granted waivers for their interests in firms that could be affected by the panel's recommendations; they could participate fully in the panel's deliberations. The Agency took into consideration other matters regarding Maureen Finnegan, M.D., John Kirkpatrick, M.D., Karen Rue, Andrew Schmidt, M.D., and Michael Yaszemski, M.D., Ph.D., who reported interests in firms at issue but in matters not related to the day's agenda; they could participate fully in the panel's deliberations. Mr. Demian then asked the panel members to introduce themselves.

Panel Chair Michael Yaszemski, M.D., Ph.D., introduced himself and stated that the purpose of the meeting was to make recommendations on a PMA for a growth factor soaked in collagen sponge to treat tibial fractures. He noted that the panel members present constitute a quorum.

OPEN PUBLIC HEARING

No comments were made.

SPONSOR PRESENTATION

Owen Fields, Ph.D., worldwide regulatory affairs, Wyeth Research, described InductOs, a combination product consisting of an absorbable collagen sponge (ACS) soaked with recombinant human bone morphogenetic protein-2 (rhBMP-2), an osteoinductive protein. The sponge is manufactured by Integra Life Sciences and is also marketed as the Helistat sponge; rhBMP-2 is produced by Wyeth. Only the 1.5 mg/mL dosage of rhBMP-2 is proposed for marketing.

The preclinical research involved both device and drug studies. The combination product is biochemically identical for all uses. Because the FDA approved the INFUSE spine fusion product in July 2002 and information on the manufacture of rhBMP-2 and ACS has therefore already been provided to the panel, Dr. Fields did not focus on the preclinical safety program.

Rod Riedel, Ph.D., project management, Wyeth, summarized preclinical studies demonstrating bone induction in animals. Bone induction is unique to rhBMP-2. The ability of the protein to induce bone growth de novo provides the rationale for replacing standard therapy components (e.g., bone graft) or augmenting standard therapy. Preclinical safety findings found that no toxicity or systemic effects occurred; the drug is eliminated rapidly, and low systemic exposure occurs. The preclinical research both in animal models and in vitro demonstrates that the protein accelerates healing and establishes the safety profile of the device. The research supports the clinical use of InductOs in fracture treatment.

Alex Valentin, M.D., senior director, clinical R&D, Wyeth, presented results of the pivotal clinical study, which was a prospective, randomized, multicenter study involving 450 patients with open tibia fractures, 300 of whom were treated with the device. Dr. Valentin presented information on the study rationale, design, patient characteristics, results, and conclusions. He noted that open fractures involve the tibia more often than any other bone.

The pivotal study was single blind and designed to use stratified randomization; patients were not aware of which treatment group they were in. Both the treating physician and an independent radiological panel assessed outcomes using the same x-ray films. Union was defined as at least three or four cortices being breached on two or more views. Patients were followed for 12 months in 7 visits. Patients receiving prophylactic bone grafting were excluded because the treatment is rarely prescribed; therefore, should the product be approved, patients requiring

prophylactic bone grafting should be excluded. Investigators could use reamed or unreamed nails because there is no international consensus on which nails work best. Patients randomized to the experimental group received standard care plus the device in either an 0.75 mg/mL or a 1.5 mg/mL dose. The study demonstrated the superiority of the 1.5 mg/mL dose compared with the control group.

Dr. Valentin noted that the study is as standardized as possible given the nature of orthopedic trauma. No consensus exists on the definition of delayed union or healing; for the study, assessment of delayed union and healing was based on the prospective collection of signs and symptoms supporting the investigators' assessment.

Dr. Valentin described the patient demographics and fracture characteristics and stated that the similarities across groups were sufficient to permit pooling the data. Patients tended to be white men in their early 30s; nearly two-thirds of the patients had been involved in motor vehicle accidents, and about half were smokers. Compared with the control group, patients in the 1.5 mg/mL experimental group experienced significantly less treatment failure and fewer interventions. Significantly more patients in the 1.5 mg/mL group experienced healing by 26 weeks without secondary intervention to promote healing, as assessed by the investigator. In addition, patients in the 1.5 mg/mL group experienced accelerated fracture union at 26 weeks, as assessed by the radiology panel.

In response to FDA concerns that the study subjects were too heterogeneous, Dr. Valentin stated that no one category influenced study results and that the homogenous efficacy results justify pooling the data. Although FDA had expressed concern that the assessments of outcomes were biased, investigator assessments of treatment success or failure were corroborated

by patient symptoms, reintervention decisions were made at comparable timepoints across treatment groups, and healed patients did not regress.

Even though the patients in the study experienced comparatively high incidence of infections, the rate was comparable across the three groups and was most common in the patients with the most severe fractures, suggesting that rhBMP-2 did not increase the likelihood of infection. Hardware failures were more likely in the control and low-dose groups and were most common for patients receiving unreamed nails. No ectopic calcifications were found to be due to rhBMP-2. The incidence of heterotopic calcifications was low, although patients receiving the InductOs device experienced a slightly higher, not statistically significant rate. In addition, antibodies to rhBMP-2 were found in a small number of patients.

Marc Swiontkowski, M.D., University of Minnesota Medical School, Minneapolis, presented information on the clinical relevance of the study findings to the U.S. trauma population. Open tibia fractures are associated with a high rate of delayed union requiring secondary procedures. The results of the study are applicable to other, less severe fractures. He pointed out that the pivotal study had a measurable endpoint, was clinically meaningful, and was consistent with other trials. The study was conducted using the highest quality design available and demonstrated that the InductOs device is safe and accelerates healing in long bones.

FDA PRESENTATION

Aric Kaiser, M.S., PMA lead panel reviewer, described the device, reviewed the indications for use, and presented the submission history of the device. He noted that this is the sponsor's second submission; the sponsor has submitted a reanalysis of the original data to address earlier deficiencies. FDA is reviewing only the tibia data presented as a result of the clinical trial. Some

outstanding preclinical issues include the effects of rhBMP-2 on tumor promotion and fetal development.

Barbara Buch, M.D., CDRH, presented FDA's clinical review. She noted several issues concerning the clinical study that related to confounding variables, patient assessment, endpoints, and data analysis. Many variables affect fracture healing, the most striking of which is the nail insertion technique. Statistically significant differences were found between the control group and the experimental groups. Patients in the experimental groups were more likely to receive reamed nails; the difference may be clinically significant. Also, the group receiving the larger dose of rhBMP-2 had a slightly higher proportion of large-diameter, unlocked nails. Fracture types were considered collectively, rather than by fracture type; however, different types of fractures were dealt with differently. Also, the protocol allowed for less than a full sponge to be inserted, so not all patients received same dose; it is unclear how that might have affected the rate of healing. Although no statistically significant differences were found between large and small centers, they handle fractures differently and have different experience levels. Diverse cultural expectations existed across sites. Half the patients were enrolled in two countries (South Africa and Germany), and several sites had fewer than 10 patients. Patient characteristics differed from country to country, and the types of fractures treated in each country differed. The ability to extrapolate to U.S. populations was not clearly demonstrated.

Dr. Buch also listed various problems with the clinical assessment methods. No standardized assessment methods were implemented. The criteria for radiographic union for independent radiology panel are problematic. A definition of healed fracture was set forth, but it was not clear whether all three of the definition's criteria had to be met. The difference between delayed union and delayed healing is unclear. Also, it is not clear how the decision to undertake

secondary intervention was made. Finally, use of prophylactic bone graft may be helpful, but the sponsor excluded those patients. Although treatment groups received a substance to enhance bone healing, the control group did not—so was the control group treated comparably?

The primary endpoint had four subgroups, but a combined clinical and radiological endpoint (CCRE) was used, an analytical method that has not been validated. The CCRE was composed of subjective assessment plus objective assessment—that is, two dissimilar assessments, rather than two radiological assessments. Patients with secondary interventions were treated differently; they were paired with the results of independent radiology review at the time the decision to intervene was made. Treatment of missing data was inconsistent.

With regard to effectiveness, Dr. Buch noted that 6 months may not have been an appropriate endpoint because the differences between the investigators' and the radiology panel's assessments were so marked. The experimental and control groups showed little difference in time to fracture healing, as determined by the independent radiology panel.

Concerning safety, Dr. Buch pointed out that the rates of authentic antibody responses are higher than in other studies of this type of device. Twelve patients (6 percent) had an authentic immune response. The contribution of the trauma setting to the response and its clinical significance is not known. Liver and pancreas function tests showed higher levels of amylase and higher rates of hypomagnesemia in the experimental group than in the control group, findings of concern because a patient in a previous study of rhBMP-2 was diagnosed with pancreatic cancer. Infection rates seem a bit higher than in other studies.

Chang S. Lao, Ph.D., Division of Biostatistics, OSB, presented the FDA statistical review. One problem is the poolability of the data due to the heterogeneity among the centers; the sponsor's approach to analyzing the data is not valid. Survival analysis is required because

patients were censored or lost to follow up. In addition, it is not clear that the study could be reproduced because so many subjective judgments were used in determining patient outcomes.

PANEL DISCUSSION

Sanjiv Naidu, M.D., summarized the preclinical results. He described several studies using animal models and concluded that not all the studies demonstrated the effectiveness of rhBMP-2 to promote fracture healing. The results are equivocal. His “gut feeling,” however, is that the protein does enhance bone formation, but its clinical usefulness will be hard to demonstrate.

Maureen Finnegan, M.D., presented the panel’s clinical review. She reviewed the problems with the pivotal study design. Many patients received rods at first intervention, which is not the standard of care in the United States. Why did several very capable countries contribute only a few patients to the study? Was the study protocol not the standard of care in those countries? Did the centers see few open tibial fractures? If so, she would be concerned about their experience. She highlighted several differences between the control group and the experimental group (e.g., the control group had a larger number of unreamed nails and more patients with multiple fractures). Dr. Finnegan noted that the sponsor stated that the ACS needs to touch both ends of the fracture, which suggests that osteoconduction is the primary method of fracture healing. CCRE has no validation associated with it.

The biggest concern is that BMP antibodies were found among patients in the experimental group. No one understands what this finding means. The safety and efficacy of repeat use of the device are unknown. In sum, one cannot conclude anything from the pivotal study. The product shows borderline efficacy, but it seems safe. All we know is that the material does not interfere with healing.

Kinley Larntz, Ph.D., provided the panel statistical review. He stated that the sponsor's study does show an effect of rhBMP-2. Nevertheless, there is a question about the poolability of the data. In the statistical analysis, one must take into account random effects. If one conducts the analysis using a logit scale, which the sponsors tried to do, one would find insufficient data to conduct the analysis. Dr. Larntz would use a Bayesian analysis, which would reveal a significant effect of rhBMP-2 with respect to the number of reinterventions.

An important aspect of the sponsor's study is the censoring of data. The censoring with respect to radiological assessment or investigator assessment was considerable. For example, in radiological assessment, the rate of censoring at 150 days was 37 percent for the control group, meaning that there were no data for radiological assessment for 37 percent of those patients after 150 days. For the group receiving 1.5 mg/mL doses of rhBMP-2, the rate was 16 percent. The differences in censoring rates between the groups is of concern. If we make adjustments for censoring, there is no difference between the three groups in terms of radiological assessment. If you adjust for the differences between reamed and unreamed nails, it reduces the effect of the 1.5 mg/mL dose on secondary interventions.

OPEN PUBLIC HEARING

Bill Christianson, Orthopedic Surgical Manufacturer's Association, urged the panel to focus on the safety and effectiveness of InductOs. He reviewed the meanings of "safe" and "effective" and noted that the regulations and the law clearly state that the standard is "reasonable." The standard is not proof beyond shadow of doubt.

***Question 1:** Discuss the adequacy of data from experience treating acute open tibia fractures stabilized with IM nails to support a more general indication.*

The panel generally agreed with the sponsor that the indication should cover long-bone fractures in general, not just tibia fractures. However, panel members noted that the data may or may not support use in open tibia fractures and that it may be problematic to extrapolate to other fractures. More data are needed.

Question 2: *Discuss the impact of the following on the ability of the study to collect clinically valid data: (1) definition of standard of care in view of the multiple confounding factors; (2) clinical relevance of the rate of secondary interventions required to promote healing as a primary endpoint; and (3) reliability of interpretation of terms “union,” “healing,” “delayed union,” and “delayed healing” at various sites.*

The panel concurred that the standard of care was acceptable with some caveats: The local standard of care may have affected the choice of nail, and other differences may have affected the infection rate. In a global sense, however, the standard of care was met. Other noninvasive means may have been appropriate additions. Also, because some patients did not get the full sponge, there was variation in the rhBMP-2 dose received. The panel also concurred that the study had flaws in the definitions of union and healing and that the investigators demonstrated enthusiasm for study material.

Question 3: *Accounting for trial design, resulting data, and statistical analyses, discuss the adequacy of effectiveness in terms of (1) the decrease in number of secondary interventions required to promote fracture healing and (2) accelerated fracture healing determined by fracture healing at 6 months assessed by investigator and fracture union assessed by independent radiologist.*

The panel concurred that the device is generally effective, but the study data are equivocal.

Question 4: *Accounting for trial design and resulting data, discuss whether or not the sponsor has provided a reasonable assurance of device safety in view of the rate of authentic antibody response to rhBMP-2 and to bovine Type I collagen, rate of hardware failure, rate of infection, and rate of abnormal liver function for lab values.*

The panel agreed that the rate of hardware failure and infections is not a major concern. Several panel members commented on the nature of the antibody response—it may warrant looking at patients over time. Trauma may have an effect. Also, panel members expressed concern about

the relation of the antibody response to tumors. The panel generally concurred that the device is safe.

OPEN PUBLIC HEARING

No comments were made

VOTE

Mr. Demian read the voting instructions. The panel voted 6-1 to approve the PMA with the following conditions:

1. Use of the device should be limited to penetrating fractures that, in the opinion of the treating surgeon, represent fractures that would benefit from device.
2. Users should be educated about the device.
3. The labeling should state that the effects of repeated use are unknown with respect to antibodies.
4. Two studies on antibodies should be conducted—a natural history study and an animal or human study.
5. Postmarket surveillance should be conducted to clarify the issue of accelerated fracture healing.
6. The sponsor and FDA should work out the statistics concerning the presence of fibula fracture fixation and its effect on endpoints.

When asked to explain the reasons for their votes, panel members responded that the device is safe, but its effectiveness is only weakly demonstrated, if at all. Many felt that the sponsor could produce a better study. The panel member voting not to approve the device did so because the preclinical data were mixed, the clinical data had too many confounders, and the x-ray data were not complete; the healing criteria were varied.

GENERAL DISCUSSION ON SPINAL DEVICES

OPEN PUBLIC HEARING

Dr. Brenda Seidman, Seidman Toxicology Services, said that FDA's proposal to require in vivo particle injection studies for spine devices is problematic. She raised numerous questions concerning the appropriateness of such a requirement.

Diane Johnson, vice president, Regulatory Affairs, Spinal Dynamics Corporation, requested that the panel consider FDA's position on durability testing and its relationship to wear-particle testing. She asked the panel to consider whether simulation testing for disc devices should reflect activities of daily living or maximum loads and motions, as sometimes suggested by FDA. She also asked the panel to consider whether particle size distribution and quantity should be determined using a load and motion profile associated with activities of daily living. Finally, she asked whether the generation of particulate due to nonphysiologic loading necessarily invalidates animal models in terms of evaluation of the effects of particulate that is generated in the animal.

Bailey Lipscomb, Ph.D., vice president, clinical affairs, Medtronic Sofamor Danek, raised several issues. First, in clinical trials, Oswestry pain success should be based on improvement from baseline, with a goal of 20 percent improvement. Second, neurological status is just one component of success; fusion, pain, and device criteria also are important. Currently, anything affecting neurological success affects the total success rate for studies. A person could have reduced pain and many sensorimotor improvements, but if one neurological criterion is reduced, it is counted as failure. The standard is inappropriate. Third, FDA is requiring unnecessarily large clinical study sizes.

Richard Jansen, Pharm.D., vice president, regulatory and clinical affairs, Disc Dynamics, Inc., raised three issues. First, in clinical studies, the radiographic endpoint of motion on flexion and extension films should be a secondary endpoint; the main endpoint should be clinical. Patients will consider surgery successful if they have reduced pain and a return to activities of daily living. Second, the baboon model is inappropriate for disc nucleus prostheses. The disc base is too narrow, and the nucleus cavity is too small. Third, spine arthroplasty devices

should not be required to last the life of the patient. Each type of device must be considered individually.

Britt Norton, Raymedica, Inc., Bloomington, MN, spoke about preclinical testing related to nonfusion spinal devices. The tests used to evaluate prosthetic nucleus devices must recognize the fundamental differences between those devices and fusion devices in order to provide meaningful results. It is important to differentiate between tests that characterize the materials used in a device and tests that evaluate the intended function of the device. He raised issues concerning tests for compression fatigue, durability, and shear testing; migration and expulsion; creep and stress relaxation; and potential for generating wear debris.

FDA PRESENTATION

Dr. Buch presented background on the guidance document for preparation of IDEs for spinal systems. She stressed that the panel discussion is not related to approval of a specific device. The current spinal guidance focuses on fusion devices, and it needs to be updated to incorporate new devices.

Panel Reviews

John S. Kirkpatrick, M.D., provided the panel preclinical and clinical reviews. He noted that the panel is being asked to recommend tests for unknown devices even though there are no validated test methods. The panel will ask for companies to satisfy the burden of safety and effectiveness. The extensive nature of the tests are high, but compromise is required.

Devices should be tested in all anticipated loads and motions. Physiologic load needs to be considered seriously. In addition, mechanical characteristics should include device changes after testing for durability, wear, debris, plastic deformation, and other changes. Migration and expulsion tests are important, as are tests for static and dynamic shear testing, creep, and stress

relaxation. Evaluation of bone–implant interface may be important. For devices intended to preserve motion, the limits need to be defined; stability testing is the way to get at that. Failure at the extremes of motion must be characterized. Biocompatibility and toxicity testing for new materials and characterization of corrosion, wear, response to debris, and shelf life all are important. The effects of debris on the spine are unknown, and animal models provide us with the best guess at present. Radiographic measures are challenging at best. Investigators should consider making indications as refined and specific as possible.

Concerning safety, sponsors should consider the revision options after device failure. Is replacement an option, or is fusion the only option? All neurological effects should be reported.

In summary, nonfusion devices will generally require more extensive evaluations, longer in vitro testing, and longer clinical followup. FDA should set forth exhaustive guidance and adjust requirements to specific device applications if a sponsor provides an appropriate rationale.

John Doull, Ph.D., provided the panel safety review. Safety boils down to whether a device has a toxic effect or a physiologic effect. Dose affects toxicity. Much is known about silicone, asbestos, and solid-state tumorigenesis. As one reduces the size of the particle, one reduces the propensity of the material to induce tumor. The kind of neurological testing one might do to look for particle-size effects does not have any good animal models. Good models are needed.

Question 1: *Please comment on the currently recommended preclinical mechanical, debris, or wear testing to evaluate new materials, device properties/integrity and wear debris for fusion and non-fusion devices. Discuss what additional testing, if any, should be added to current testing recommendations for the following devices: stabilization devices for non-fusion; intervertebral disc/joint replacements (cervical/thoracolumbar); devices manufactured out of new materials; and intervertebral disc nucleus replacements.*

The panel concurred that the guidance adequately specifies the necessary tests. The problem is that nonfusion devices come in many types of materials and polymers. Viscoelastic testing is

required, and there are many types of tests. Because the materials are so variable, these issues should be addressed individually with the device sponsors.

Question 2: *Because of the limitations of the current testing methods and models, should devices made of new materials and/or those intended to retain motion be tested for neurotoxicity independent of the type of material or the amount of wear debris generated? If you suggest that testing be performed, please describe the testing you recommend.*

The panel concurred that wear-debris testing is important. Members proposed various types of testing, such as placing particulates in the area of intended use adjacent to the dura and testing in conventional modes. Physiologic loading to represent the most vigorous patient type, instead of maximal loading, would be appropriate. The panel concurred that it is important to place a barrage of particles at the site of implantation and study the reaction. Panel members noted that the manufacturer should not be required to produce particles of a size not produced under wear conditions and that FDA needs to be careful not to put too high a burden on manufacturers.

Question 3a: *Please discuss study designs which may be better suited to evaluate Non-Fusion spinal devices. In your discussion, please comment on enrollment criteria, patient populations, controls, success criteria and goals of the study that would be suitable for these types of non-fusion spinal devices.*

The panel concurred that it is important to look at the natural history of nonfusion devices; the devices may affect degenerative changes at adjacent segments. Investigators need to carefully consider specific indications for each device. Success should include improvement of pain and functional capacity. Patient success is multivariate construct.

Question 3b: *Devices intended to stabilize the spine yet retain functional motion are expected to have an upper limit of motion beyond which one would consider the device to be unstable and a lower limit below which one would consider the device to have inadequate motion or possibly even consider the segment to be fused. Please discuss the amount of motion and on what scale, to define a patient as a functional and clinical success (i.e., a clinically significant improvement in condition) for each of cervical, thoracic and lumbar levels for Non-Fusion spinal devices.*

The panel concurred that it was not useful to require an arbitrary level of motion. If a patient receives a nonfusion device and improves, then the device is efficacious. The panel concurred

that is important to study a wide range of diagnoses; indications for replacement versus fusion are important factors in product design. Investigators must determine whether interventions would affect adjacent deterioration. Success criteria should include pain level as well as range of motion.

ADJOURNMENT

Mr. Demian thanked the participants and adjourned the panel at 4:22 p.m.

I certify that I attended this meeting of the Orthopaedics and Rehabilitation Devices Advisory Panel on November 21, 2002, and that these minutes accurately reflect what transpired.

Hany Demian, M.S.
Executive Secretary

I approve the minutes of the November 21, 2002, meeting as recorded in this summary.

Michael Yaszemski, M.D.
Acting Chairperson

Summary prepared by
Caroline G. Polk
Polk Editorial Services
1112 Lamont St., NW
Washington, DC 20010
(202) 265-8271
cpolk@earthlink.net