The Design and Use of Animal Models for Translational Research in Bone Tissue Engineering and Regenerative Medicine

George F. Muschler, M.D.,1 Vivek P. Raut, M.S.,1 Thomas E. Patterson, Ph.D.,1 Joseph C. Wenke, Ph.D.,2 and Jeffrey O. Hollinger, D.D.S., Ph.D.3

This review provides an overview of animal models for the evaluation, comparison, and systematic optimization of tissue engineering and regenerative medicine strategies related to bone tissue. This review includes an overview of major factors that influence the rational design and selection of an animal model. A comparison is provided of the 10 mammalian species that are most commonly used in bone research, and existing guidelines and standards are discussed. This review also identifies gaps in the availability of animal models: (1) the need for assessment of the predictive value of preclinical models for relative clinical efficacy, (2) the need for models that more effectively mimic the wound healing environment and mass transport conditions in the most challenging clinical settings (e.g., bone repair involving large bone and soft tissue defects and sites of prior surgery), and (3) the need for models that allow more effective measurement and detection of cell trafficking events and ultimate cell fate during the processes of bone modeling, remodeling, and regeneration. The ongoing need for both continued innovation and refinement in animal model systems, and the need and value of more effective standardization are reinforced.

Introduction

This review is written with several goals: (1) To provide a rational paradigm of generalizable principles that will guide investigators in the choice of optimal model systems and in the design of new model systems. (2) To provide an overview of the current animal models that are available for exploring and advancing therapeutic options for bone repair and regeneration. (3) To define some of the current deficiencies in the current models or their use, and to articulate some opportunities for further development. In doing so, the authors hope to contribute to the ongoing professional and cultural dialog regarding the use of animal models, and to accelerate the rate of innovation and effective development of new and more effective clinical therapies and patient care.

Tissue engineering and regenerative medicine related to bone include a broad range of settings and approaches that seek to repair, augment, replace, or regenerate bone tissue.1,2 Formation of new tissue by bone-forming cells (osteogenic cells) is a central feature of each of these goals, and can broadly be placed into four major strategic categories of cell therapy: (1) “targeting” local stem cells and progenitors at a local site (e.g., scaffolds, bioactive factors, and biophysical methods), (2) “transplantation” of stem cells or progenitors (harvest from one site and transplantation to a new site, with or without processing to increase concentration or prevalence in a local site), (3) “homing” of stem cells and/or progenitors into a wound site from regional tissues or systemic circulations (e.g., using native bioactive factors or selective targeting using selective binding or magnetic fields), and (4) “modification” of stem cells or progenitors (e.g., in vitro expansion or adaptation with or without transient or durable genetic modification that alters or enhances performances).1 Experiments designed to assess questions and optimize conditions related to the clinical translation of each of these strategies require means of systematic control and modification of key variables associated with the implementation of these strategies. A single ideal model is not currently available in which bone tissue regeneration can or should be assessed. However, the most effective model systems for bone regeneration research and development should meet several essential criteria. (1) They should provide an environment that matches, to the greatest extent possible, the clinical and biological environment and material formulation in which the methods under assessment will be used. (2) They should provide objective and quantifiable parameters (metrics) to assess the success (quantity and quality) and functional performance of the regenerated bone tissue. (3) They should detect and

1Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio.
2United States Army Institute of Surgical Research, Fort Sam Houston, Texas.
3Department of Biomedical Engineering and Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania.
1. REPORT DATE  
01 FEB 2010

2. REPORT TYPE  
N/A

3. DATES COVERED  
-

4. TITLE AND SUBTITLE  
The design and use of animal models for translational research in bone tissue engineering and regenerative medicine

5a. CONTRACT NUMBER  
-

5b. GRANT NUMBER  
-

5c. PROGRAM ELEMENT NUMBER  
-

5d. PROJECT NUMBER  
-

5e. TASK NUMBER  
-

5f. WORK UNIT NUMBER  
-

6. AUTHOR(S)  
Muschler G. F., Raut V. P., Patterson T. E., Wenke J. C., Hollinger J. O.,

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  
United States Army Institute of Surgical Research, JBSA Fort Sam Houston, TX

8. PERFORMING ORGANIZATION REPORT NUMBER  
-

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)  
-

10. SPONSOR/MONITOR’S ACRONYM(S)  
-

11. SPONSOR/MONITOR’S REPORT NUMBER(S)  
-

12. DISTRIBUTION/AVAILABILITY STATEMENT  
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES  
-

14. ABSTRACT  
-

15. SUBJECT TERMS  
-

16. SECURITY CLASSIFICATION OF:  

<table>
<thead>
<tr>
<th>a REPORT</th>
<th>b ABSTRACT</th>
<th>c THIS PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
</tr>
</tbody>
</table>

17. LIMITATION OF ABSTRACT  
UU

18. NUMBER OF PAGES  
27

19a. NAME OF RESPONSIBLE PERSON  
-

Standard Form 298 (Rev. 8-98)  
Prescribed by ANSI Std Z39-18
predict clinically relevant differences in biological performance between methods (i.e., the relative rate and frequency of success).

Bone as a Composite Tissue—Implications in Defining Metrics for Assessment

Rational approaches for animal models to assess bone regeneration demand an understanding of the nature of bone as a complex and highly organized dynamic and heterogeneous connective tissue that provides biological function in many domains: mechanical load bearing and force transmission, immunologic function (leukogenesis and lymphogenesis), mass transport (erythrogenesis), coagulation (thrombogenesis), energy storage (fat deposition), and calcium homeostasis (Fig. 1).

The defining feature of bone is an extracellular collagen based matrix (osteoid) that is mineralized with a unique carbonated hydroxyapatite. This mineralized bone matrix provides the unique biomechanical properties of elastic resistance to deformation in compression and bending that allows bone to transmit the mechanical forces enabling mobility, mastication, and respiration, and to shield critical organs from trauma (e.g., the eyes and brain). As a result, almost all studies of bone tissue engineering utilize an assessment of volume, distribution, and density of mineralized matrix as a primary outcome measure (e.g., radiographs, computed tomography, or X-ray absorptiometry). MicroCT imaging in particular has revolutionized the utility of many in vivo models. MicroCT may provide quantitative information regarding three-dimensional pattern and distribution of mineralization at the macromolecular level of bone structure, enabling assessment of trabecular number, thickness, connectivity, and mineral density, which are correlated strongly with mechanical properties and the status of modeling and remodeling within a region.

However, bone is more than just a mineralized tissue. Assessments of other parameters are necessary to provide a full quantitative description of the quality of bone. Mineralized bone matrix, despite its unique features and elegance, represents only a fraction of the overall volume of skeletal tissues. In normal cancellous trabecular bone, mineralized bone matrix generally represents less than 25% of the total tissue volume. The majority of the volume of bone tissue consists of a highly variable mixture of fat, fibrous tissue, and hematopoietic tissues that contain an equally variable, but generally rich plexus of arteries, veins, and venous sinusoids.

Histological assessment is generally the key secondary metric of outcome. Histology is necessary to characterize the pattern and distribution of cellularity in the mineralized bone tissue compartment and/or the kinetics of bone formation and mineralization (osteocytes, osteoblasts, lining cells, and osteoclasts). Histology is also necessary to characterize the nature of tissue within the nonmineralized intertrabecular regions of bone with respect to hematopoietic cells, fat, fibrous tissue, vascularity, and the status of any local inflammatory response or residual implant materials. Immunohistochemistry and in situ hybridization can further refine characterization of cell phenotypes based on local matrix, cell surface markers, intracellular markers, and gene expression.

Chemical characterization of the composition of both organic matrix (collagen and noncollagenous proteins) and/or mineral is sometimes needed to distinguish bone matrix and true biological mineralization (carbonated microcrystalline hydroxyapatite) from scar or regions of precipitation of amorphous calcium phosphate salts. However, although important as biological parameters, in the absence of known or suspected genetic, hormonal, or nutritional abnormalities, or biochemical defects, chemical and mineral analyses of new bone tissue are not generally key metrics.

Mechanical performance is, in many settings, the ultimate test of the success of bone tissue engineering. Although the volume, distribution, and density of mineralized bone matrix may be a strong predictor of the mechanical properties of new bone tissue, this correlation is not perfect, particularly in the setting of ongoing remodeling in or around an implant. As a result, mechanical testing, at a whole bone or local bone level, remains a highly relevant parameter to assess the outcome of new bone tissue (i.e., modulus, strength, ductility, and fatigue resistance). Mechanical integration of new bone tissue with adjacent bone and soft tissues or with durable implant materials (e.g., endoprosthetics) is also highly relevant to many clinical applications. Functional bone provides broad surfaces and focal points of attachment for specialized tendons, ligaments, muscle cartilage, teeth, and nails (i.e., claws or hoofs). Similarly, the capacity of a tissue-engineered construct to develop appropriate nonintegrated surfaces can be equally important in some clinical situations (e.g., gliding surfaces for specialized tendons or bursa, and surfaces that support membrane tissues lining the skull, spinal canal [dura], and chest cavity [pleural]).

The recent evolution of stem cell and progenitor cell biology has also highlighted the fact that bone and bone marrow also serve as a niche and reservoir for a heterogeneous population of stem cells and progenitors capable of contrib-
ute to the formation of all of the above tissues (not only for bone, fat, vascular cells, and hematopoietic cells, but also for cells capable of generating other tissues, such as liver and nerve). The capacity of regenerated bone to become repopulated with an appropriate population of stem cells and progenitors has not yet become a clinically relevant metric of outcome, but may be exploited more in the future. Repopulation with stem and progenitor cells may be a predictive factor of the long-term performance and remodeling potential of the new tissue. This may be particularly important to long-term clinical success in settings of regional stem cell and progenitor cell deficiency, such as radiation injury, large segmental defects, or congenital pseudoarthrosis.

**In Vivo and In Vitro Animal Models for Bone Repair**

The discipline of engineering translates fundamental knowledge in physics, chemistry, and biology into materials, devices, systems, and strategies to achieve practical benefits. Engineered solutions, almost by definition, include the systematic assessment and attempt to optimize each variable that may contribute to success or failure. Tissue engineering and its clinical outlet in regenerative medicine apply this conceptual framework to advance the repair, replacement, or regeneration of organs and tissues. As a discipline that is fundamentally grounded in living systems, the design and development of quantitative and reproducible model systems for *in vitro* or *in vivo* study is essential to the rational development of tissue engineering. The tool sets available in existing *in vitro* and *in vivo* models are, as they always have been, incomplete. There will always be a need to design and develop new animal models to address new questions or to better resolve old questions in a more rigorous and standardized fashion.

*In vitro* assays using cell culture preparations have found increasing use in the discovery of fundamental biological mechanisms (e.g., transcription, translation, signaling events and processes, and cell cycle regulation), characterization of the effects of materials on isolated cell function, and for screening large numbers of compounds for biological activity, toxicity, or immunogenicity. Bioreactor systems have begun to expand the application of these systems of discovery to include three-dimensional constructs and some multicellular composites. However, *in vitro* systems are inevitably limited in their capacity to recreate the complex *in vivo* environment, and are ultimately incapable of predicting *in vitro* or clinical performance in many settings, particularly in settings of tissue engineering and regeneration. Moreover, new and promising therapies such a systemic doses of parathyroid hormone cannot be evaluated in *in vitro* models.

*In vivo* models (i.e., animal models) are required when *in vitro* systems cannot provide a reproducible approximation of the real-life *in vivo* or clinical setting. Some obvious and common examples include the kinetics of delivery and distribution of drugs or bioactive factors; the biocompatibility and degradation properties of implant materials; the effect of intervention on the local secretion or release of cytokines and the flux and stability of *in vivo* cytokine gradients (which define the heterogeneous and locally varying chemical environment of wounds); the patterns and distribution of gradients associated with the mass transport of oxygen and other nutrients; the survival of cells and tissue after transplantation; regional distribution cellular kinetics (proliferation, migration, differentiation, matrix synthesis, apoptosis, and necrosis) through tissue compartments; migration of cells through vascular systems (both immune/inflammatory response and stem cell or progenitor cell homing); the multicellular processes associated with budding angiogenesis and revascularization to reestablish the flow of blood in a tissue; and the heterogeneous environment of mechanical stimuli associated with muscle contraction and mechanical deformation through extrinsic forces or intrinsic movement (e.g., flexion and extension of a limb or spine) and vascular pulsation. Although selected aspects of these processes can be studied *in vitro*, none of these environments or processes can be created reliably *in vitro* outside of very narrow and uniform conditions. Moreover, *in vivo*, all of these systems vary simultaneously across time and space, which cannot be reduced or isolated during embryonic development and morphogenesis, during disease initiation and progression, during a response to injury or therapeutic intervention, or during aging.

**Domains of Use for Animal Models**

Animal models play a critical role in many domains of study in medicine and biology. These can generally be categorized in three domains: (1) fundamental discovery, (2) feasibility and bioactivity testing, and (3) clinical modeling and efficacy prediction.

Questions related to fundamental discovery involve discovery and characterization of biological mechanisms, addressing the questions framed in the context of, “How do things work?” and “What is the origin and mechanism of morphogenesis or disease?” These questions are inevitably addressed in highly defined animal models (both genetically and immunologically) that provide a high level of reproducibility. Fundamental discovery research demands animal systems where reagents (e.g., antibodies and probes) are readily available and where turnover of individuals (breeding) and the occurrence of biological events (e.g., morphogenesis, disease development, injury repair, and aging) are rapid. Therefore, the study of animals with short development and life cycles (days to months) provides the best opportunities for detection and observation of the process under study. Common choices for fundamental research include *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (nematode worm), *Danio rerio* (zebra fish), and many varieties of *Mus musculus* (mouse), and *Rattus norvegicus* (rat). Technical methods to design and develop transgenic animals have greatly enabled the rigorous systemic and even focal conditional modulation of specific genes alone and in combination in these species, and have dramatically advanced the rate of fundamental biological discovery in the past decade. The sequencing of these genomes and the general homology of many biological processes across large evolutionary gaps has broadened the relevance of discoveries across species. However, in the case of studies on bone tissue, only bony vertebrae, fish, and higher vertebrates are specifically relevant.

Feasibility and bioactivity testing takes one step toward clinical translation, by asking, “Is this mechanism a possible therapeutic target?” and “Will modulation of this mechanism using this specific set of local or systemic means influence the development or progression of disease or repair without unacceptable toxicity?” Feasibility and bioactivity testing
generally includes assessment of biocompatibility, toxicity, screening for adverse reactions, and the capability of delivering a given device or agent in a manner in which it has a measurable and desirable effect.

The same models that are used to test feasibility are often used as screening systems for quality control by confirming bioactivity or performance of existing reagents or products and determine whether the methods of fabrication or processing are changing over time. Models used in this domain are generally designed to deliver the therapy or agent in a manner that closely approximates methods of delivery in a clinical setting (oral, IM, IV, tissue infusion, or open implantation), and to examine specific tissues or fluids that are most likely to be sensitive to change (e.g., blood, urine, and tissue). In the setting of an implant, the anatomic location and tissue at the site of implantation will generally match the type of tissue, if not the exact location, in which therapeutic use is expected. However, in the case of bone formation or repair, both subcutaneous, epimuscular, intramuscular sites and periosseous and intraosseous sites of implantation (both cranial and long bone) have been used.

Feasibility testing is almost always done in small mammals and specifically in mice and rats. Many standard models for the assessment of implant materials have been codified in standard protocol documents published by major standards and regulatory organizations, including the American Society for Testing and Materials (ASTM), the International Standards Organization (ISO), the United States Food and Drug Administration (FDA), and the European Commission. However, for many relevant strategies and biological processes (e.g., delivery of bioactive factors, cell transplantation, cell homing, and migration), existing models may not provide optimal or generally accepted tools for feasibility testing. As a result, considerable opportunity remains for the development of models for rapid in vivo feasibility assessment based on the function of cells and bioactive factors.

Like model systems used for fundamental discovery, models used for feasibility and bioactivity testing benefit from the use of inbred strains of mice and rats, where outcomes can be determined after relatively short periods of observation (days to weeks), where reagents (e.g., antibodies and probes) are readily available, and where the variation in radiographic, imaging, histological, or biochemical outcome between individuals is small (reducing the number of animals needed to achieve a statistically significant assessment). When evaluating implant materials, models enabling the assessment of multiple samples in a single animal subject are also preferred, both providing control for variation between animals and limiting the number of animals needed. All of these factors help limit the time, expense, and subject burden needed to complete a given body of research.

In some settings, feasibility studies must be advanced into a larger animal, such as a rabbit, dog, sheep, or goat. However, these settings are generally limited to just a few situations: (1) when the surgical procedure involved cannot be performed reproducibly in a smaller animal (e.g., spinal fusion procedures or tendon/rotator cuff repair procedures), (2) when the size of the implant or device under study exceeds the volume capacity of a smaller animal (e.g., assessment of an implant over 1 cm, depending on the tissue site), and, similarly, (3) when there is the need to model effects of large diffusion distances or void volumes, as in the case of mass transport limitations (e.g., diffusion of oxygen) in the survival of transplanted cells.

Clinical modeling and efficacy prediction is also generally referred to as “preclinical animal testing.” In this setting, animals are used to test the questions, “Is this therapy, material, or method optimized for performance in the clinical setting in which it is intended?” or “Will this method perform as well or better than an existing standard therapy?” Animal models used to address questions in this domain are designed to create an environment that is as close as possible to the clinical setting in which a therapy will be used. Therefore, key criteria of these models and studies are to (1) deliver the therapy in the manner in which it will be delivered in a clinical setting; (2) utilize a site of delivery and/or assessment site that is as closely matched as possible to the setting(s) in which it will be used (anatomic site, anatomic size, and local tissue characteristics — presence of fat, muscle, scar, and vascular bed), (3) utilize surgical techniques (when applicable) that match or are most analogous to clinical methods; (4) utilize an animal that provides a metabolic background and physiological responsiveness comparable to humans, and (5) utilize a formulation of active agent that has the same composition, dose, release, retention, and degradation properties as the formulation that will be used clinically.

These criteria can be applied rigorously in settings evaluating synthetic and nonimmunogenic biomaterials or reagents. However, in some settings biological factors force some modifications in the last criterion. For example, when evaluating some bioactive active agents (e.g., growth factors such as bone morphogenetic proteins [BMPs]), large species-specific variations in efficacy and dose responsiveness may be present. As a result, preclinical assessment in this setting can be complicated by the need to identify and utilize an animal species in which responsiveness is comparable to humans or to utilize a species-specific formulation that is modified to create an effect that is comparable in the model being used to the bioactivity expected in the clinical setting.

In the setting where allograft tissue-derived materials are being assessed, immunological barriers make implantation of human tissues into animal subjects (a xenograft) an inappropriate model. As a result, assessment of allograft tissues in an animal model requires not that human allograft tissues be used, but that the methods of tissue procurement and processing are as closely matched to clinical methods used on human tissue as possible. Fidelity between the materials being assessed is therefore not based on standardized formulation and composition, but rather is based on the validateable presumption that composition of bone matrix from humans and other large animals will be altered similarly when identical processing methods are used.

The preclinical assessment of strategies utilizing autograft or allograft cells is similarly limited by immunological barriers that make transplantation of human cells into animals as xenografts inappropriate as a model for clinical application. As a result, specific methods of cell harvest and processing become the variable that can and should be modeled to the clinical standard. This relies on the validateable assumption that the starting composition and intrinsic biological potential of tissue-derived cell populations from large animal species (including progenitors and stem cells) is comparable to cell populations that may be harvested from human subjects and that identical
processing steps will result in comparable changes in composition and relative biological potential and performance.

Special challenges are presented, however, when the clinical cell processing or validation methods can only be applied to human cells, due to the need to use species-specific reagents (e.g., antibodies that do not cross react with the corresponding functional homolog in an appropriate animal or where no nonhuman homolog exists). In this case, preclinical studies might require use of a surrogate reagent or process that provides comparable processing effect to the missing reagent. Thereby, the variable that can be assessed in the preclinical setting becomes a process with a particular outcome (e.g., a comparable change in composition of the cell population), rather than exact method of processing that would be used clinically to achieve this end result.

Far more substantial challenges and risks are presented when cell processing involves in vitro culture expansion, rather than direct transplantation or immediate intraoperative processing. In this setting, species differences in optimal culture conditions (serum and growth factors requirements/responsiveness) confound the opportunity to directly apply validated human methods or to refine in vitro systems for a given species with the presumption that they would provide comparable effects on the selective proliferation of subsets of proliferative cells or on differentiation or long-term biological potential to those achieved by highly refined systems developed for human clinical application. Preclinical systems must also be established to assess the long-term fate of culture-expanded cells, and to detect and limit the potential for selection or development of clones with undesirable or potentially harmful biological properties during periods of rapid expansion outside of the normal systems of systemic immune surveillance.32,33

Some useful general guidelines have been published by the FDA relevant to the design of preclinical studies to assess cell therapy strategies.34 Additional proactive attention to the development of a more rigorous system of nomenclature for the stem cell and progenitor cell populations that are the target of cell therapy development and of better defined standards and models for in vitro and in vivo are needed, and will greatly enable and accelerate the development of safe and effective cell therapy strategies.

Biological Principles and Technical Considerations in Choosing an Animal Model

The design and selection of animal models for clinical translation is intimately related to the therapeutic modality that is being tested. A review of potential modalities of therapy is beyond the scope of this paper. Several recent reviews of this topic are available.35–39 Overall, the tools and materials that may be applied to these settings include scaffold materials (e.g., allograft bone, synthetic polymers, and calcium phosphate ceramics), cell sources (autogenous bone, autogenous marrow, local tissue, and culture-expanded autograft of allograft cells with or without genetic modification), bioactive factors (e.g., BMPs, vascular endothelial growth factors, platelet-derived growth factors, epidermal growth factor, and parathyroid hormone [PTH]), and mechanical/biophysical modulators (cast/brace immobilization; internal or external fixation; and electrical, magnetic, mechanical, or ultrasound stimulation).

A useful conceptual paradigm is to recognize that virtually all therapeutic modalities can be categorized based on one or more of the following five domains of biological activity or function1:

Osteogenesis  the process by which osteogenic stem cells and progenitors create new bone tissue through processes of homing, activation, proliferation, migration, differentiation, and survival. This includes all modes of cellular therapy, that is, the homing or harvest, processing, and transplantation of osteogenic cells.

Osteoconduction  that is, scaffold functions preservation of a defined tissue volume in a manner that prevents the encroachment of adjacent tissues and that promotes the attachment, migration, distribution, and survival of osteogenic cells within a defined tissue volume.

Osteoinduction  traditionally defined as the process by which soluble or matrix-bound signals interact with local cells (progenitors or nonprogenitors) to initiate a cascade of cellular events that change the fate of local uncommitted stem cells or progenitors toward an osteoblastic phenotype.

Mass transport, that is, the general process of flux and exchange of chemical modalities through tissue by convection (physical and mass movement along pressure gradients in vascular channels or in extracellular fluid spaces exposed to deformation via movement or external deformation), diffusion (passive migration of chemical species along concentration gradients of submicron to millimeter scale), or active pumps (usually acting only across membranes at short distances). Mass transport modalities, would therefore include modalities to preserve of enhance vascular transport (including vascularized tissue grafts) and means to enhance revascularization through angiogenesis.

Biophysical effects  generally grouping modalities of mechanical forces (strain, shear, compression, and pressure), temperature, and electrical and/or magnetic field effects.

In general, every living system in which bone exists (fish to mammals) will be influenced by these variables. However, variables of physical size, anatomy, and species-specific histology and physiology make some models better suited for investigation of individual domains than others. For example, in the domain of osteoconductivity, screening of a given scaffold as a bulk material with respect to variables of biocompatibility, surface osteoconduction, toxicity, and degradation rate can be performed in small defects and using solid discs, screws, or rods using bone defects of 2–5 mm in maximal dimension and as thin as 1 mm in size (generally in mice, rats, and rabbits). In contrast, evaluating a porous scaffold material, in a setting where mass transport is the primary variable and the question under study is the capacity of a porous scaffold to induce the ingrowth of new blood vessels as a delivery system for a bioactive factor enabling the survival cells across a gradient of profound hypoxia within a wound environment, requires work in a larger defect measuring 1–5 cm in maximal dimension but no thinner than 5–10 mm (generally dog, sheep, and goat).

In the domain of osteogenesis, rodents have provided an effective setting for screening for evidence of stem cell and
progenitor cell homing to sites of bone repair, evaluation of the differential effects of marrow-derived and periosteal-derived cell populations, and in screening for the effects of cell transplantation across major and minor histocompatibility barriers using nonvascularized, vascularized, and even whole limb transplantation. These studies are greatly enabled by means for marking and tracking cell progeny from different sources over time (discussed in more detail below).

However, rodent models begin to fail when questions evolve toward clinical application. In addressing questions of cell and tissue harvest methods and methods for processing autogenous cells for transplantation, rodent models are not well suited. The anatomy and composition of autogenous bone and marrow in rodents provides a dearth of cancellous trabecular bone (rodent long bone are primarily cylinders with little metaphyseal cancellous bone except immediately adjacent to the physsis). As a result, autograft controls in rodent inevitably represent a mixture of bone marrow with minced cortical bone and periosteum; this composition is not comparable to the purely cancellous bone grafts used clinically and available in dog, goat, and sheep models. Rodent-derived marrow cells are also intrinsically different in biological performance than marrow from large animals and humans. When marrow is irrigated from rodent long bone (the most common method of harvest) the fraction of osteogenic colony forming units (CFU) is only 10% of all fibroblastic CFU. In contrast, in human marrow aspirates, over 90% of fibroblastic CFU colonies are osteogenic (i.e., can be classified as osteogenic CFU). Moreover, in addressing questions related to the limitations of mass transport and hypoxia on the survival of transplanted cells in a clinically relevant setting, it is necessary to evaluate defects in which diffusion distances are clinically relevant, in general 3-5 mm or more from the nearest vascular bed (i.e., minimum diameter or thickness of 6-10 mm). Defects in this range can only be accommodated in large animals (dogs, sheep, goats, and pigs). A more detailed discussion of animal selection in the domain of translational models is addressed below.

New cell therapy strategies that involve the local recruiting or homing of autogenous or engineered cells to a site of injury or therapeutic interventions have recently increased the need to rigorously define and track the fate of cells that may be implanted, injected, or homed from systemic circulation. In this domain, mouse models have gained renewed importance. Tracking of cells from a defined source and their progeny has been most effective using green fluorescence protein (or engineered variants) as a constitutive or conditionally expressed marker. Mouse calvarial defects and fibular fracture repair models have been used effectively to define the process of progenitor cell homing from systemic circulation.

### Practical Issues in Design and Selection of Animal Models for Clinical Translation and Preclinical Testing

A myriad of variables must be rationally defined to insure that the assessment achieved using the model will be most relevant to the clinical domain of interest. Subsequently, as many variables as possible must be controlled to minimize random effects and to make variation in outcome as sensitive as possible to the effects of test variables and to minimize the number of animal subjects needed to achieve statistical power. Among the many variables are animal subspecies; animal subject inclusion criteria (age, sex, partum, and gonadal and estrus/lactation status); nutritional status; diet; exercise/activity regime; animal temperament (often a function of species and vendor); anesthetic, pharmacologic, and pain management; the setting of bone repair/regeneration (fracture, osteotomy, and defect); anatomic location (bone and site); use of fixation (bracing, internal, and external); the timepoints for assessment; the relevant endpoints of outcomes to assess; and the means for assessment (radiographic, histological, biochemical/molecular, and mechanical). Further, within the domain of a repair/regeneration setting are further variables, such as local tissue environment (preservation or removal of periosteum or local marrow, and the character of adjacent tissue, i.e., fat, muscle, and fibrous tissue or scar). In the setting of fracture, fracture location and pattern must be defined using a consistent mechanism of injury. In the case of osteotomy, the location and geometry of the osteotomy as well as the means by which it is created must be defined. Large differences have been shown between surgical techniques (corticotomy vs. osteotomy), use of sharp versus dull instruments, reaming with or without a tourniquet, and the management of periosteum and local soft tissues. The importance of defining and following precise and reproducible and standardized protocols that define each of these variables cannot be understated, not only to enable objective assessments within a given lab or project, but also to enable objective comparisons between laboratories.

Given the large opportunity for variation among so many variables, the selection of the most appropriate model to address a given scientific question has been principally a question of judgment and opinion, and will remain so for at least the next few years. However, the goals and principles of defining the conditions in each model are generally the same: (1) limit variation from animal to animal as much as possible, (2) define methods for assessment that are sensitive to change in the most important outcome (e.g., bone formation, bone bridging, remodeling, and revascularization), (3) utilize treatment conditions/materials that most closely match the material or treatment that will be used clinically, and (4) utilize a biological tissue bed that most closely matches the tissue bed that will be present in a clinical setting. Every setting in which materials are tested seeks to achieve these goals, but no model environment can maximize all of these variables simultaneously. Small uniform defects in animals of uniform age, sex, and genetic background, which maximize the first two goals, fall miserably short on the third, matching the clinical conditions of a wound site. This is particularly true regarding the variables of mass transport environment (e.g., oxygen diffusion/hypoxia, concentrations/gradient of signaling molecules, and revascularization rate) and progenitor cell kinetics (e.g., progenitor concentration/distribution, survival of local or transplanted progenitors, or progenitor cell homing or migration events), which are highly dependent on wound size. At the other extreme, the gold standard of a human clinical trial will, by definition, exactly match the biological tissue bed in the clinical setting, and the clinical material (Goals 3 and 4) fails miserably to limit variation (Goal 1) and enables only limited radiographic and...
histological assessment (Goal 2). Some large sources of clinical variation can be excluded by setting age, sex, disease, body mass index, site, and other exclusion criteria, as is common in clinical studies. However, clinical trials of bone repair and regeneration inevitably include subjects with large variation in systemic biology, including genetic background (including receptor phenotypes),62–64 immunological sensitivities, endocrine status, nutritional status, comorbidities, and pharmacological effects. They are also associated with large differences in local biology at a given site of bone repair, including defect size and shape, local vascularity, the health and composition of local tissue (marrow, bone fat, muscle, and fascia), the presence of scar (prior surgery, trauma, and radiation), or biophysical environment (loading, shear, and strain).

The scope of this review precludes detailed discussion of the decision-making process linked to the many variables involved in a given model. The following sections provide a very short overview, drawing heavily on information that is summarized well in several recent reviews.36,65–72

Selection of Animal Species

A wide variety of mammalian species are used in bone research. Figure 2 illustrates the differences in size, shape, and internal architecture of bone among 10 mammalian species that are commonly used for bone repair and regeneration research. Figure 3 compares cortical and trabecular microstructures of their distal femur samples. In general, preclinical translational testing is performed in large skeletally mature animals, rather than rodents or rabbits. Dog, goat, sheep, and pig are the most utilized species.65 The use of these animals provides a skeletal biological environment that is not undergoing ongoing growth or modeling (a lifelong process in rodents, which do not fuse their growth plates). The use of large quadrupeds also provides the opportunity to utilize autogenous cancellous bone as a control or gold standard for contemporary bone grafting. Cancellous bone can be harvested from the proximal humerus (the largest bone in quadrupeds) using a small cortical window in quantities and in a consistency that closely matches cancellous hematopoietic bone and marrow from the human iliac crest. Large animals also allow the use of fracture, osteotomy, and defect sites as well as methods of internal and external fixation that more closely match those used in clinical settings.

Nonhuman primates (NHPs), African green monkeys (Cercopithecus aethiops), Rhesus macaque (Macaca mulatta), and baboon (Papio hamadryas) have been used in studies of bone repair. The use of NHPs adds substantially to the cost of research and is associated with some cultural and ethics questions. The primary rationale for NHPs is the opportunity to match the genetic background and therefore the biological responsiveness of the model as closely as possible to humans. It has been assumed that this will provide the most predictive model for immunological and biological response, particularly in the setting of delivery of human peptide growth factors such as BMPs. However, the variation in response between primate species can be as great as or greater than variation between humans and quadrupeds. Baboons, for example, are reported to be hypersensitive to BMPs.20,21,73,74 One disadvantage of NHP models, other than baboon, is that they are anatomically much smaller in size than other large animals. The adult Rhesus weighs only 6.5 12 kg (male), and has a femur measuring 16.5 cm in length and 1.25 cm in mid shaft diameter (as compared to a coon hound dog measuring 20 cm in length and 2.5 3 cm midshaft diameter). This compromises the opportunity to match the defect volumes and fixation methods available in other models.

Therefore, although these primate models are conceptually appealing, there is at this time no definitive consensus nor compelling need to propose that the preclinical assessment of new technologies for bone repair and regeneration must pass through a NHP model. In the future, in vitro screening of biomaterials and active agents and direct comparison of the in vitro response of human subject animal cells to a bioactive agent may become an appropriate step to confirm that a given animal model provides an appropriate biological background for assessment.

O’Loughlin et al.65 recently summarized the reported use of animal models for study of fracture repair in long bones based on articles published in six orthopaedic journals over a 10-year period. These papers included fractures without

![FIG. 2. Comparison of femoral size and anatomy among animal species commonly used for bone regeneration research. Anterior view of left femurs of adult animals (from left to right): mouse, rat, rabbit, dog, goat, sheep, pig, South African monkey, rhesus, baboon, and human. Color images available online at www.liebertonline.com/ten.](image-url)
gaps, noncritical-sized gap defects (which would generally heal spontaneously during the lifetime of the animal), and critical defects, which will not heal without intervention. The relative use in this domain in order of frequency was rat 38%, rabbit 19%, mouse 13%, sheep 11%, dog 9%, goat 4%, and other 4%. Closed models of fracture repair involving a blunt guillotine mechanism\textsuperscript{79} have been used in approximately 40% of studies involving the rat and mouse. However, almost all studies in larger animals utilize open defects. Internal or external fixation has been used with roughly equal frequency. Roughly 20% of research studies using mice, rats, and rabbits reported no use of fixation other than external immobilization (generally ulna and fibula models). However, immobilization alone was used in few models in larger animals.

Utilization of animal models in the spine has not been reviewed in similar detail. Mouse and rat models are generally avoided because of their small anatomic size. The rabbit intertransverse posterior fusion model without internal fixation\textsuperscript{76–83} has been a common starting point. Rabbit intervertebral fusion is not utilized. The rabbit intervertebral space has a marked tendency for rapid degeneration and even spontaneous fusion following relatively minor trauma to the annulus.\textsuperscript{84} Canine models for posterior fusion have been described with\textsuperscript{85–92} and without\textsuperscript{85,87} internal fixation. Goats\textsuperscript{96–103} and pigs\textsuperscript{104–110} models allow more opportunity for internal fixation anteriorly or posteriorly. However, it must be noted the horizontal loading of the lumbar spine in quadrupeds (all animal models including NHPs) does not reproduce upright posture of bipedal humans with respect to axial compression and rotational loading in the human lumbar spine. This incongruity in the mechanical loading of spine models can be less profound in the cervical spine of sheep and goats, where a relatively upright posture have been exploited effectively as models of human cervical bone grafting.\textsuperscript{111–115} Fusion models in NHPs are generally limited to the lumbar spine and performed both anteriorly\textsuperscript{116–119} and posteriorly,\textsuperscript{120,121} and with\textsuperscript{116,122} and without\textsuperscript{120} internal fixation.

Evaluation of biomaterials and surgical strategies for craniofacial applications has appropriately utilized many animal models, including calvarial defects,\textsuperscript{71,123–145} maxillary sinus augmentation,\textsuperscript{146–155} alveolar ridge augmentation,\textsuperscript{156–171} and mandibular defects.\textsuperscript{172–182} In addition to defining the genetic background, the choice of animal species also establishes boundary conditions to the work that can be performed in many other dimensions, including the range of possible defect sizes, options for reagents (cells, growth factors, antibodies, and probes), parameters of defect fixation and mechanics, and even the possible duration of experiments (dependent upon life span). Table 1 provides an overview of the key features of various animal groups and species.\textsuperscript{183–189}

Mimicking the underlying bone biology of the human clinical setting is one of the principle goals of selecting a given animal. The biology of these species differs in several important respects. For example, rodents continue to model (i.e., grow and reshape) their skeleton throughout their lives. Growth plates remain open. In addition, the trabecular bone content of rodent bone is limited, even in metaphysis of long bones. Moreover, Haversian remodeling by tunneling osteoclasts does not occur in rodents. All of these factors are significant negatives when considering rodents as appropriate models of human bone biology. In contrast, rabbits (i.e., lagomorphs no longer classified as rodents) have both larger and relatively more trabecular bone mass that exhibits Haversian remodeling. However, rabbits have very fatty (oily) marrow that is distinctly different in physical properties than human marrow, making them an undesirable choice as a model for autogenous bone and marrow harvesting, processing, or transplantation. Rabbits also are notoriously sensitive to glucocorticoid stimulation, which results in profound intraosseous fatty hypertrophy and secondary avascular necrosis of bone.\textsuperscript{190–192}

Bone composition and biology of dog, sheep, goat, and pig are very similar to those of humans. Detailed reviews of the robust literature comparing bone structure, composition, and biology in these animals\textsuperscript{86,90} identify modest differences in cancellous and cortical bone density at various sites, differences in response to ovariectomy and dietary calcium restriction, magnitude of sexual dimorphism, the age at which peak bone mass is achieved, and the rate and extent to which Haversian remodeling replaces plexiform lamellar bone in the

<table>
<thead>
<tr>
<th>Mouse</th>
<th>Rat</th>
<th>Rabbit</th>
<th>Dog</th>
<th>Goat</th>
<th>Sheep</th>
<th>Pig</th>
<th>South African Monkey</th>
<th>Rhesus</th>
<th>Baboon</th>
<th>Human</th>
</tr>
</thead>
</table>

FIG. 3. Comparison of cortical and trabecular microstructures. MicroCT scans of distal femurs are depicted. (A) The pattern and distribution of cancellous bone in the distal femur and metaphysis. MicroCT resolution: 45 μm for all bones except for mouse and rat, which are scanned at 27 μm. (B) Cross section of the cortical diaphysis along the plane indicated by the red line in (A). (C) Cancellous bone from the medial femoral condyle, illustrating trabecular architecture. Dimensions of the trabecular cube: 5×5×5 mm (0.125 cm³). Color images available online at www.liebertonline.com/ten.
<table>
<thead>
<tr>
<th></th>
<th>Mouse</th>
<th>Rat</th>
<th>Rabbit</th>
<th>Dog</th>
<th>Goat</th>
<th>Sheep</th>
<th>Pig</th>
<th>South African monkey</th>
<th>Rhesus macaques</th>
<th>Hamadryas baboon</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common name or type</strong></td>
<td><em>Nude mouse/house mouse</em></td>
<td>Rat/Sprague Dawley rat</td>
<td>New Zealand rabbit</td>
<td>Beagle, coonhound, or mongrel dog</td>
<td>Spanish boat goat/South African goat</td>
<td>Domestic sheep</td>
<td>Domestic pig</td>
<td>Mona monkey</td>
<td>Rhesus macaques</td>
<td>Hamadryas baboon</td>
<td>Human</td>
</tr>
<tr>
<td><strong>Scientific name</strong></td>
<td><em>Mus musculus</em></td>
<td><em>Rattus norvegicus/rattus</em></td>
<td><em>Oryctolagus cuniculus</em></td>
<td><em>Canis familiaris</em></td>
<td><em>Capra hircus</em></td>
<td><em>Ovis aries</em></td>
<td><em>Sus scrofa</em></td>
<td><em>Cercopithecus mona</em></td>
<td><em>Macaca nemestrina</em></td>
<td><em>Papio hamadryas</em></td>
<td><em>Homo sapien</em></td>
</tr>
<tr>
<td><strong>Adult weight</strong></td>
<td>12–30 g</td>
<td>70-300 g</td>
<td>1.50–2.50 kg</td>
<td>Beagle: 8.2–16 kg Coon-hound: 32–40 kg</td>
<td>45 kg avg</td>
<td>110 kg</td>
<td>50–350 kg</td>
<td>Male: 5 kg Female: 4 kg</td>
<td>Male: 6.5–12 kg Female: 5.5 kg avg</td>
<td>Male: 21.5 kg Female: 9.4 kg</td>
<td>76–83 kg</td>
</tr>
<tr>
<td><strong>Life span</strong></td>
<td>2 years</td>
<td>4 years</td>
<td>9 years</td>
<td>20 years</td>
<td>15 years</td>
<td>10–12 years</td>
<td>10 years</td>
<td>20–30 years</td>
<td>37 years</td>
<td>50–80 years</td>
<td>16–18 years</td>
</tr>
<tr>
<td><strong>Age of skeletal maturity</strong></td>
<td>Continuous growth</td>
<td>Continuous growth</td>
<td>10–11 months</td>
<td>12–18 months</td>
<td>3 years</td>
<td>15–18 months</td>
<td>12–14 months</td>
<td>Not available</td>
<td>7 years</td>
<td>5 years</td>
<td>12–18 years</td>
</tr>
<tr>
<td><strong>Age of sexual maturity</strong></td>
<td>5–7 weeks</td>
<td>3–5 months</td>
<td>8 months</td>
<td>6–12 months</td>
<td>5–10 months</td>
<td>6–8 months</td>
<td>9 months</td>
<td>Male: 4.8–6.8 years Female: 4.3 years</td>
<td>Highly sexually dimorphic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sexual dimorphism only in size</td>
<td>Highly sexually dimorphic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
cortex of long bone. However, these points of variation do not provide evidence that any one of these animals provides a better match for human bone biology than the other. As a result, all of these animals provide excellent options. As a result, considerations of expense, local access, temperament, ease of husbandry, personal or institutional experience, the need or desire to reproduce an established precedent, and unique anatomic considerations become the legitimate drivers of decisions between these options. Pigs tend to be least used,\(^{66}\) likely due to difficult temperament, anatomically short and thick long bones, and large size. The short bones do not allow for large gap defects because adequate fixation is difficult to achieve. Minipigs are more manageable, but may also be less desirable due to concern that these genetic dwarfs may be less biologically similar to humans than common domestic animals. Dogs are used most in North America both due to a rich background of prior experience, ease of husbandry, and accessibility. Purpose-bred beagles (a relatively small dog) or coon hounds (a large dog) provide reproducibility and accessibility with minimal ethical implications. Purpose-bred animals are useful, but may also be less desirable due to concern that these genetic dwarfs may be less biologically similar to humans than common domestic animals. Dogs are used most in North America both due to a rich background of prior experience, ease of husbandry, and accessibility. Purpose-bred beagles (a relatively small dog) or coon hounds (a large dog) provide reproducibility not offered by more heterogeneous mongrel dogs.

A robust literature characterizing canine models supports continued preferential use of dogs in many settings, particularly in bone marrow biology and transplantation immunology.\(^{193–206}\) The use of dogs is being challenged, however, by social concerns regarding the dog as a companion animal. Sheep are long boned, and allow use of many systems of internal fixation used in human clinical practice. Sheep are readily available and docile, but can be frail and less resilient, particularly when not in a group. Goats, similar in size and skeletal anatomy to sheep, have a thinner coat and less muscular bulk, and are more tolerant of stress and isolation. Goats are also becoming more available in the United States, making them an increasingly viable alternative for bone research.

NHPs are sometimes held up as the model of choice for prediction of clinical performance for bone repair and regeneration methods. This supposition is based on arguments of evolutionary proximity and similar structural anatomic, endocrine, digestive, and dietary features and immunology. However, it is difficult to justify a view that the use of NHPs should be proposed as a gold standard for work in bone repair and regeneration. Defect size is an important limitation in the African Green Monkey and Rhesus. Of the three common NHP species, only the baboon approaches the skeletal size range that humans share with large canines, sheep, and goats. NHP species exhibit profound sexual dimorphism (males larger than females) that is far greater than humans. Moreover, there is heterogeneity in the biological response among NHPs that diminishes the argument that evolutionary proximity necessarily equates to improved clinical prediction. For example, in preclinical studies of BMPs the baboon has been reported to be hypersensitive to BMPs and to require doses of BMPs substantially lower than Rhesus to induce the same effect.\(^{20,21,73,74}\) NHPs are very expensive and present husbandry challenges that preclude models involving intensive wound assessment, management, or immobilization. In addition, although NHPs do not have companion animal status, arguments of perceived evolutionary affiliation raise social and ethics questions related to increasing their use in research. Most important, however, methods that work well in NHPs tend to work well in other large animals and vice versa. As a result, although exceptions must be considered, there is a lack of direct evidence that NHP models are systematically or consistently superior to other large animal model as a means of objective assessment comparison between implanted devices, cell transplantation strategies, or growth factor delivery systems, with respect to biological mechanism or prediction of either relative or absolute clinical performance than other large animal models.

**Systemic Factors—Age, Sex, Endocrine, and Pharmacological Status**

Systemic variables are well-established variables having important effects on the outcome of bone repair and regeneration. This has implications both on the clinical relevance and on the sensitivity and responsiveness of a given model. Many studies have demonstrated effects of endocrine status, including estrogens, androgens, thyroid hormone, PTH, PTH-related protein (PTHrP), and glucocorticoids, in settings of intrinsic cyclic variation (e.g., estrous cycle), surgical ablation, genotypic variation, or pharmacological administration.\(^{70,207}\) Therefore, the genetic background and endocrine status become important to define and control.

Age has also been identified as a key systemic variable influencing the character and outcome of bone repair and regeneration, though these studies have generally been limited to date to small animal models.\(^{207,208}\) In large animals, expense considerations tend to drive research to be performed using young animals soon after they reach skeletal maturity, due to the increased cost of raising purpose-bred animals to an advanced age. The sensitivity of young animals as models for detection of the therapeutic efficacy or toxicity associated with some therapies is limited, because young animals and humans may not express deficiencies in stem cell biology, angiogenesis, growth factor synthesis, or signal transduction that may become manifest later in life. Therefore, near exclusive assessment of bone repair and regeneration strategies in young healthy animals may result in a collective distortion of outcome. On one hand, the base level of performance of a given strategy is likely to be overestimated relative to the clinical setting. On the other hand, exclusive use of young animals may also underestimate the incremental effect of adding adjuvant methods directed at age-related deficiencies and even the characterization of dose response effects and the detection of toxicity associated with some therapies. An increased use of aged animals is one of the possible means to enhance to refine existing animal models. Aged rodents can be purchased from certified vendors and the National Institute on Aging (www.nia.nih.gov/Research Information/Scientific Resources/). In addition, purpose-bred animals retired from breeding stock (usually females) are a potential cost-effective source of aged large animals.

**Local Biological Factors—Defect Size and Type**

Fractures in animals and humans generally heal with a high frequency of success. Failure of repair generally occurs only in settings where one or more complicating factors are present, including (but not limited to) fracture or gap in bone (caused by either tissue loss or distraction maintained by internal or external fixation), local tissue loss (especially periosteal tissue or local intramedullary marrow), interposition of soft tissue (fat, muscle, and tendon), local osteonecrosis or soft tissue necrosis, local infection, extensive scarring, excessive mechanical mo-
tion, biological compromise resulting from dietary deficiency, endocrinopathy (especially parathyroid, thyroid, and adrenal abnormalities), pharmacological effects (e.g., non-steroidal anti-inflammatory drug (NSAIDs), chemotherapy, and glucocorticoid use), and radiological effects (local radiation therapy).209

In the design of a clinically relevant animal model system, each of these variables can be and have been used as means to “elevate the bar” and increase the biological challenge to bone tissue regeneration and the extent that treatment would be required to achieve bone repair. In the clinical setting, of course, no one factor is present. It is always a combination of these factors that determines the biological challenge and the risk of treatment failure. However, of all these possible variables that can be used, a defect or gap in bone is both most easily standardized. Defects can generally be divided into cylindrical defects, which can be reproducibly created using a trephine or drill, and segmental defects, created using a transactional saw. Hemicylindrical defects have been used, but less frequently.212

Critical-sized segmental bone defects have become the mainstay of translational models to evaluate and compare bone repair strategies. In 2009 the ASTM published the Standard Guide for Preclinical In Vivo Evaluation in Critical Sized Segmental Bone Defects (F2721-09),210 which serves as an excellent resource. Critical defects are defined as “a defect that will not heal without intervention.” The femur and tibial diaphysis tend to be used most often, but require internal fixation (rod, plate, or external fixation). Radius, ulna, and metatarsal models have also been used, and can be used without fixation in many settings. The actual defect size that is necessary to preclude healing without intervention varies between species and anatomic sites. However, as a rule of thumb, a segmental defect in the diaphysis of a long bone that is 1.5–2.0 times the diameter of the bone involved will result in a nonhealing defect.

Segmental defects are readily created using standardized methods. Although these models are considered to be highly reproducible, there are many sources of minor variation that can influence outcome, and can contribute to systematic differences between the biological environment of one model compared to another, and differences with respect to the clinical environment. Some of these sources of variation are enumerated below.

The biological environment or envelope of fresh or acute segmental bone defects is generally bounded at the ends by normal cortical diaphyseal bone (including periosteum, endosteum, and bone marrow), and surrounded by normal muscle, mixed with some fat and remnants of fibrous muscle insertion sites to the excised bone. Their uniform dimensions and defined anatomic location make them readily assessable with respect to new bone formation and revascularization using radiographic or histological methods. Mechanical outcome can be assessed using standardized measurement of torque and bending stiffness and loading to failure. Sampling for assessment of biochemical or molecular markers can also be anatomically and geometrically defined.

Cylindrical defects in the frontal or parietal bone of mouse,132,211,212 rat,213–215 and rabbit134,216,217 have in many ways been the primary screening tools or point of entry for the assessment of many biomaterials. Like segmental defects, cylindrical defects have the value of being reproducible both in size and local anatomy. They also generally provide a stable mechanical environment that does not require internal fixation, thereby removing effects of stress shielding and any compromise of radiographic assessment associated with metallic implants used for internal fixation.

In the setting of cranial defects, the tissue envelope that defines the biological setting of bone repair is considered to be relatively reproducible. In every defect, the vascular dura floating on intracranial fluid is always at the base. Cortical bone sandwiching cancellous bone and marrow is distributed circumferentially around the rim at a generally uniform thickness. Cranial sutures are generally avoided. This allows highly reproducible positioning of an implant that is precisely matched to the defect, usually without internal fixation. The tissue overlying the implant site is dependent on surgical technique and may be defined by retained periosteum, Galen’s fascia, subcutaneous fat, or dermis and skin. In the setting of an ungrafted defect, dura and overlying soft tissue will encroach into the defect and compromise bone regeneration, making this model very responsive to the performance of scaffold materials, even if they simply provide the mechanical integrity needed to resist soft tissue encroachment and preserve the space for bone formation.

In the setting of cylindrical defects in long bones, the tissue envelope that defines the biological setting of bone repair is again highly defined.218–229 The composition of the defect opening generally includes periosteum, cortical bone, and endosteum. After closure, the opening of the defect is bounded by overlying muscle or fat. The base of the defect may be cancellous bone and marrow, cortical bone, or muscle/fat, depending on defect length and orientation within the long bone. The tissue quality within the defect particularly, periosteum, cancellous bone, and bone marrow will depend strongly on the anatomic site, species, and age. In general, defects in the metaphysis of young large animals will be comprised of highly vascular hematopoietic marrow. However, marrow in defects in the diaphysis will generally be fatty or yellow marrow, which is still vascular, but far less cellular. Advancing age in humans and animals will also increase the ratio of fatty to hematopoietic marrow, changing the biological environment of the graft site. Soft tissue encroachment does not compromise bone formation in most cylindrical defects, so mechanical resistance to encroachment is not necessary for success in these models. Regardless, cylindrical defects can be very sensitive to detecting differences in efficacy between osteoconductive scaffold materials as well as the effects of bioactive factors and cell transplantation.

The uniform radial geometry of cylindrical defects also provides the potential for control and modeling of biological gradients relevant to bone regeneration. Due to the pattern of diffusion and vascular ingrowth that must be oriented from the periphery to the center of the defect, cylindrical defects are particularly well suited for time-oriented studies of mass transport (e.g., the influence of diffusion, cellular metabolic demand, local oxygen delivery or generation, or gradients in the delivery of bioactive factors); the degradation rate of implants; cell transplantation (survival and migration); and variables modulating the rate of bone or tissue ingrowth or vascular invasion into an implant.

The canine femoral multidefect model (CFMD) introduced by Takigami et al.218 is one example of a refined cylindrical defect model in a large animal that enables evaluation of variables of metabolic demand, mass transport, and
Revascularization in large defects (Fig. 4). In contrast to most large animal models in which only one graft site can be created per animal, the CFMD model provides four 10-mm-diameter and 15-mm-long defects per animal, enabling intrasubject comparison of materials or therapeutic strategies. These defects are of sufficient size that the biological environment in the interior of the defect is characterized by profound hypoxia, a key feature of clinical defects that is a limiting factor in cell survival and transplantation. Both bone formation and revascularization within the defect occur through a process of ingrowth that has a radially oriented “outside in” pattern that can be readily measured and characterized using microCT or histological methods. The importance of achieving a defect size in which diffusion becomes a limiting factor is particularly important when working with cells that rely upon diffusion for survival and with delivery of bioactive factors that must be distributed by a diffusion gradient. Diffusion rapidly becomes inefficient as defect size increases. Further, as defect size increases, the defect surface area increases proportionally to the radius of the defect; however, defect volume increases proportionally by the square of the radius. The biological environment in the CFMD model crosses this threshold (Fig. 4A), and is markedly different from that in small defects in which the distance from the edge of the defect to the center of the defect is less than 3 mm (e.g., mouse and rat defects). In small defects, although oxygen tension may fall, cell survival is possible throughout the defect, supported by diffusion of oxygen and nutrients from the periphery. Similarly, in small defects the diffusion gradients of implanted bioactive factors are short, enabling dosing nearer the physiological range.

Need for Standardization of Animal Models

The rational selection and use of animal models in all domains of use, from discovery, to feasibility testing, to
preclinical modeling, depends on reproducibility, sensitivity, specificity, biomechanical, and biological relevance. An animal model must provide the same or similar results in repeated experiments in the same lab and in identical experiments in different labs (reproducibility). Further, such models must enable detectable differences in outcome based on change in a particular test variable (sensitivity), as well as allow experimental design that isolates single independent variables for assessment (specificity). Defining reproducible methods for surgical and nonsurgical management and reproducible, sensitive, and specific means of documenting outcome using quantitative metrics (e.g., physical assessment, imaging, biochemistry, histology, and biomechanics) is absolutely essential in this process. However, this is not an easy process. Living systems, particularly large animals, are complex and prone to substantial random variation. Moreover, outcome in each subject is subject to variation from multiple sources, which include but are not limited to the effect of the operator implementing an intervention (skills, and minor variation in technique), the animal (species, subspecies, genetics/epigenetics, temperament, age, sex, and systemic and local effects), experimental variables (e.g., material, dose, composition, activity, and interactions between these), environmental effects (facility protocols, and effects of housing, diet, and activity), and observer (skills, practice, and bias) or detection/observation method (sensitivity and specificity), as well as random effects.

At the level of individual laboratories, standardization is accomplished through rigorous documentation of standard operating procedures and through the use of hands-on training methods to pass operating techniques from one operator to the next. Standardization among laboratories is most often accomplished through transfer of protocols in peer-reviewed publications. However, published materials are often limited in communicating new or refined surgical as well as processing techniques. As a result, sharing of formal protocols including detailed descriptions of technique and staged photographs is often preferable. In many cases, important details are best communicated through direct personal interaction, and even a site visit to an experienced laboratory is necessary to enable hands-on experience as an operator or observer.

Overall, these methods for standardization within and between laboratories are highly variable. This fact likely leads to a high degree of variation within the reported literature, and within and between laboratories, even when nominally using the same model.

On a national scale, the ASTM has become increasingly active in developing guidelines and standards related to the use of animal models. The ASTM Subcommittee on Medical and Surgical Materials and Devices (F-04) has published several useful and generalizable standards that impact on the field of bone tissue engineering (Table 2). Of these, ASTM F2721-09, updated in 2009, provides Standard Guide for Preclinical In Vivo Evaluation in Critical Sized Segmental Bone Defects, is of particular relevance to this review.

The ISO-published standard ISO-10993 also provides a series of standards for preclinical assessment of medical devices (Table 3). In particular, ISO 10993-6, "International Standards for Biological Evaluation of Medical Devices," provides excellent guidelines for the selection criteria for choosing an appropriate animal model for evaluation of biocompatibility.

<table>
<thead>
<tr>
<th>ASTM standards</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>F04.04</td>
<td>Division IV TEMPs</td>
</tr>
<tr>
<td>F04.05</td>
<td>Computer-assisted orthopedic surgical systems</td>
</tr>
<tr>
<td>F04.11</td>
<td>Polymeric materials</td>
</tr>
<tr>
<td>F04.12</td>
<td>Metallurgical materials</td>
</tr>
<tr>
<td>F04.13</td>
<td>Ceramic materials</td>
</tr>
<tr>
<td>F04.15</td>
<td>Material test methods</td>
</tr>
<tr>
<td>F04.16</td>
<td>Biocompatibility test methods</td>
</tr>
<tr>
<td>F04.21</td>
<td>Osteosynthesis</td>
</tr>
<tr>
<td>F04.22</td>
<td>Arthroplasty</td>
</tr>
<tr>
<td>F04.25</td>
<td>Spinal devices</td>
</tr>
<tr>
<td>F04.33</td>
<td>Medical/surgical instruments</td>
</tr>
<tr>
<td>F04.41</td>
<td>Classification and terminology for TEMPs</td>
</tr>
<tr>
<td>F04.42</td>
<td>Biomaterials and biomolecules for TEMPs</td>
</tr>
<tr>
<td>F04.43</td>
<td>Cells and tissue-engineered constructs for TEMPs</td>
</tr>
<tr>
<td>F04.44</td>
<td>Assessment for TEMPs</td>
</tr>
<tr>
<td>F04.46</td>
<td>Cell signaling</td>
</tr>
</tbody>
</table>

ASTM, American Society for Testing and Materials; TEMPs, tissue engineered medical products.

The Missing Links Between Preclinical and Clinical Performance

The development, use, and interpretation of data from preclinical animal models remain a complex challenge and an imperfect science. Available animal models are deficient in predicting clinical performance in several domains: underestimation of variation in clinical response, overestimation of performance, and insensitivity to incremental improvement. Underestimation of clinical variation in outcome is, of course, predictable and expected. Limiting variation is desirable, if not necessary, feature of a useful model when testing feasibility and variation in formulation. However, this can be a liability in clinical translation. The local biological environment in clinical settings in which bone tissue engineering and regenerative medicine strategies are applied are far more variable than in available animal models. The systemic biology in the clinical setting is also far more variable than in animal models both in age and in the addition of comorbid conditions (e.g., diabetes, tobacco use, nutritional status, obesity, osteoporosis, and estrogen/androgen status). As a result, both absolute and relative performance in the clinic may often fall away from preclinical predictions when these extra factors are added, alone, and in combination.

Overestimation of clinical performance is also a predictable development, although it was not necessarily true at the time when the concept of critical-sized defect models was introduced. Critical-sized defects were not designed as a model for the clinical setting. In contrast, they were designed simply to be just large enough that they would not heal without intervention, but small enough to be sensitive to minor improvement. Subsequently, critical-sized defect models became the target against which essentially all preclinical materials were developed and optimized. It is predictable that over time,
through incremental improvement, available materials and methods have now been refined to the point where even large defects in large animals can be healed at a high rate of success. However, as discussed in the previous paragraph, when materials are then used clinically in comparable defects but in settings of systemic and local biology that were not part of the previous optimization scheme, performance will predictably fall. Several different technology strategies are now available that achieve virtually 100% success rates in large animal defect models in preclinical studies, but demonstrate substantial failure rates and, in some cases, clinical complications that were not predicted by preclinical models.

Insensitivity to incremental improvement is also a predictable outcome of systematic optimization of materials for a limited set of models, be the critical-sized defects or spine fusion models. As available materials and strategies are improved, results improve, narrowing the remaining window for detecting further benefit. This situation has now developed even in the setting of rigorous 5-cm defects in large animal models, where several treatment strategies now demonstrate near 100% success. This creates an urgent need for more rigorous and more biologically relevant preclinical models. This challenge is discussed further under the section Gaps and Opportunities, below.

The final missing link between preclinical prediction of performance and true clinical performance is the fact that the true clinical use and performance of virtually all current products and bone tissue engineering strategies is not known. Actual clinical results of bone grafting procedures are reported in only a small percentage of cases. Data for some materials in some clinical settings are available from randomized trials, though this is limited almost exclusively to trials that are required by the FDA or another agency. Most often, outcome is reported only in relatively small retrospective case series of 10–100 patients in a limited field of application and is based on the experience of a small number of surgeons at elite medical centers. However, regardless of setting, these reports are inevitably prone to at least two sources of bias, selective inquiry (choosing to study what is already suspected to be a good series) and publication bias (choosing or being able to publish only positive results), and as a result the systematic underreporting of adverse outcomes.

Currently, all but one of the FDA-approved products for bone repair have been approved by the 510-k mechanism. This mechanism of approval does not require large-scale premarketing clinical studies or systematic data collection or systematic reporting of clinical outcomes during the postmarketing period. These current conditions make it very difficult to determine the settings in which the current products are used, the way in which they are being used, or to provide an objective correlation between performance in preclinical studies and clinical performance.

This deficiency in our current knowledge of the actual utilization and objective or relative performance across the many products now marketed for bone repair and regeneration and other applications has become more apparent as both the number of competing product options and expectation for performance have increased. This recognition is catalyzing an important national discussion, legislation, and investments in research funding to develop capabilities for a national registry of medical devices and biomedical implants. Despite these limitations and caveats, preclinical assessment of new therapy options using robust and well-defined animal models remains a central pillar in the armamentarium of tissue engineers seeking to advance clinical practice.

### Gaps and Opportunities to Improve upon Existing Animal Models

Despite these advantages of critical-sized defect models, there are obviously some major limitations. Understanding
the effects of aging on the clinical efficacy has already been mentioned as one major gap. The second gap is the dependency of existing models on critical-sized defects, in which only defect size is used to “elevate the bar,” without including other systemic and local biological factors that are known to have profound clinical relevance and effects.

The defects used in virtually all current animal models involve grafting into defects in which the biological environment is characterized by a clean, freshly cut vascular bone and a relatively untraumatized and well-vascularized soft tissue envelope. This biological environment may be found in a defect created by the excision of a low-grade malignant tumor (e.g., low-grade chondrosarcoma) in which no systemic chemotherapy or local radiation therapy will be used, but will be true in very few other settings. In contrast, the local biological environment in the vast majority of clinical defects is inevitably characterized not just by a bone defect, but also by combined deficits of local periosteum, soft tissue (muscle and fat), compromise of vascularity in local tissues, and often the effects of local changes in tissue quality and composition resulting from local disease or prior injury and scarring.

Some investigators using some models do attempt to add these local features to the model. Some examples include extending the excision, elevation, or trauma to periosteum beyond the defect by a fixed distance, and irrigating or reaming bone marrow tissue from the bone ends at the edge of the defect. The effect of these insults may improve modeling of the clinical setting in at least three ways. First, they may reduce the local concentration and prevalence of osteogenic progenitors by removing or injuring tissues in which these populations reside. Second, the procedures may simply compromise local vascularity to bone and soft tissue around the defect site, creating a more profound challenge to mass transport of oxygen and other nutrients in the site. Finally, the extra trauma may increase the amount of local necrosis, inflammation, bleeding, and subsequent scarring. Other parameters can be imposed, such as excision of local muscle or fat, cautery, and freezing of local tissue. All of these factors are more difficult to perform in a reproducible fashion than cutting out a bone defect, but the combination of these additionally local biological factors will likely be necessary to “raise the bar” of existing defect models to both enhance the clinical relevance and specificity of animal models, and to enhance the sensitivity of these models in detecting further incremental improvements in efficacy.

Conclusions

Bone is a complex composite organ composed of a unique mineralized matrix that functions in a highly integrated fashion with vascular, hematopoietic, and adipose tissues. As a result, the scientific questions related to bone health, pathology, and disease and regeneration demand the use of a broad range of animal models, and the continued need for innovation in new model systems to advance the field of regenerative medicine. Opportunities for advancement of animal models are identified in at least four domains: (1) the continued development of models to address new questions and relevant variables and outcome parameters (e.g., cell survival, cell homing and cell state transition events, mass transport, in vivo cytokine delivery kinetics, and defined genetic or signaling modification); (2) application of new methods of quantitative analysis (e.g., imaging modalities, histology or biophysical to elucidate features of existing models that have not been generally accessible); (3) improved standardization, validation, and rigorous use of models in competitive analyses that are intended to provide predictive value with respect to relative clinical performance; and (4) the design and implementation of models that better represent the complex environment in which clinical bone regeneration procedures are often performed and most often fail (larger regions characterized by bone and tissue loss, scarring, compromised vascularity, and a functional deficiency of local stem cell and progenitor cell populations).

Acknowledgments

The authors would like to acknowledge Lyman Jellema and the Cleveland Museum of Natural History for assistance with femur samples and associated discussion. The authors would also like to thank Richard Rozic, Amit Vasanji, and the Cleveland Clinic Musculoskeletal Imaging Core/Image Processing and Analysis Center for assistance and guidance regarding the microCT scanning of intact femurs.

Disclosure Statement

The authors have no commercial associations or competing financial interests in connection with this article.

References

76. Boden, S.D., Moskovitz, P.A., Morone, M.A., and Tor ibiake, Y. Video assisted lateral intertransverse process
arthrodesis. Validation of a new minimally invasive lumbar spinal fusion technique in the rabbit and nonhuman pri mate (rhesus) models. Spine 21, 2689, 1996.


137. Acarturk, T.O., and Hollinger, J.O. Commercially available demineralized bone matrix compositions to regenerate
198. Vasseur, P.B., Rodrigo, J.J., Stevenson, S., Clark, G., and Sharkey, N. Replacement of the anterior cruciate ligament...
with a bone ligament bone anterior cruciate ligament allo

L., Li, X.Q., Zika, J.M., and Goldberg, V.M. Cyclosporin A and tis
sue antigen matching in bone trans
plantation. Fibular allografts studied in the dog. Acta Or

200. Bos, G.D., Goldberg, V.M., Powell, A.E., Heiple, K.G., and Zik,
A.M. The effect of histocompatibility matching on ca

201. Goldberg, M., Luk, S.C., Greyson, N.D., and Greenaway, A. Cani

202. Goldberg, V.M., Bos, G.D., Heiple, K.G., Zika, J.M., and Pow,

ham, T.C., Torok Storb, B.J., and Thomas, E.D. Abrogation of resis
tance to and enhancement of DLA nonidentical un
related marrow grafts in lethally irradiated dogs by the ra

ham, T.C., and Thomas, E.D. Resistance to marrow grafts in


206. Zaucha, J.M., Zellmer, E., Georges, G., Little, M.T., Storb, R., Stor,
er, B., and Torok Storb, B. G CSF mobilized periph
eral blood mononuclear cells added to marrow facilitates engraftment in nonmyeoloblated canine recipients: CD3 cells are required. Biol Blood Marrow Transplant 7, 613, 2001.


208. Kloss, F.R., and Gasser, R. Bone and aging: effects on the ma


214. Itagaki, T., Honma, T., Takahashi, I., Echigo, S., and Sasa
no, Y. Quantitative analysis and localization of mRNA transcripts of type I collagen, osteocalcin, MMP 2, MMP 8, and MMP 13 during bone healing in a rat calvarial exper


218. Takigami, H., Kumagai, K., Latson, L., Togawa, D., Bauer, T., Powell, K., Butler, R.S., and Muschler, G.F. Bone for

mation following OP 1 implantation is improved by addi


222. Markel, M.D., and Sielman, E. Radiographic study of homo


226. Beniker, D., McQuillan, D., Livesey, S., Urban, R.M., Turner, T.M., Blum, B., Hughes, K., and Haggard, W.O. The use of acellular dermal matrix as a scaffold for peri


Address correspondence to:
George F. Muschler, M.D.
Department of Biomedical Engineering
Lerner Research Institute
The Cleveland Clinic Foundation
9500 Euclid Ave. (ND20)
Cleveland, OH 44195

E-mail: muschlg@ccf.org

Received: October 1, 2009
Accepted: November 5, 2009
Online Publication Date: January 7, 2010
This article has been cited by:


5. Dr. Thor Einar Friis, Dr. Sally Anne Stephenson, Prof. Yin Xiao, Dr. Jonathon Paul Whitehead, Dr. Dietmar Werner Hutmacher. A PCR-based method for isolating clones from a cDNA library in sheep. *Tissue Engineering Part C: Methods* 0:ja.. [Abstract] [Full Text PDF] [Full Text PDF with Links]


21. Yuchun Liu, Jerry K Y Chan, Swee-Hin Teoh. 2012. Review of vascularised bone tissue-engineering strategies with a focus on co-culture systems. Journal of Tissue Engineering and Regenerative Medicine n/a-n/a. [CrossRef]
25. J.-P. Boutrand. Methods and interpretation of performance studies for bone implants 271-312e. [CrossRef]
28. Roger Brooks. Clinical Aspects of the Use of Stem Cells and Biomaterials for Bone Repair and Regeneration 493-520. [CrossRef]


