
http://www.elsevier.com/locate/issn/87563282

http://www.sciencedirect.com/science/journal/87563282

http://dx.doi.org/10.1016/j.bone.2014.07.010

http://medlib.mef.hr/2330
The rational use of animal models in the evaluation of novel bone regenerative therapies

Mihaela Peric, Ivo Dumic-Cule, Danka Grcevic, Mario Matijasic, Donatella Verbanac, Lovorka Grgurevic, Vladimir Trkulja, Cedo M. Bagi, Slobodan Vukicevic

a University of Zagreb School of Medicine, Center for Translational and Clinical Research, Laboratory for Mineralized Tissues, Salata 11, Zagreb, Croatia
b Department for Intercellular communication, Salata 2, Zagreb, Croatia
c University of Zagreb School of Medicine, Department of Physiology and Immunology, Salata 3, Zagreb, Croatia
d Department of Pharmacology, Salata 11, Zagreb, Croatia
e Pfizer Inc., Global Research and Development, Global Science and Technology, 100 Eastern Point Road, Groton, CT 06340, U.S.A.

# To whom correspondence should be addressed: e-mail: vukicev@mef.hr; tel. +385 1 4566812; fax. +385 1 4566822

Keywords: bone fracture, bone healing, animal model, regenerative therapy, tools for scoring, healing parameters
Abstract

Bone has a high potential for endogenous self-repair. However, due to population aging, human diseases with impaired bone regeneration are on the rise. Current strategies to support bone healing include various biomolecules, cellular therapies, biomaterials and different combinations of these. Animal models for testing novel regenerative therapies remain the gold standard in pre-clinical phases of drug discovery and development. Despite improvements in animal experimentation, excessive poorly designed animal studies with inappropriate endpoints and inaccurate conclusions are being conducted. In this review, we discuss animal models, procedures, methods and technologies used in bone repair studies in an attempt to assist investigators in planning and performing scientifically sound experiments that respect the wellbeing of animals. In the process of designing an animal study for bone repair investigators should consider: skeletal characteristics of the selected animal species; a suitable animal model that mimics the intended clinical indication; an appropriate assessment plan with validated methods, markers, timing, endpoints and scoring systems; relevant dosing and statistically pre-justified sample sizes and evaluation methods; synchronization of the study with regulatory requirements and additional evaluations specific to cell-based approaches.

Highlights

• Animal models in bone regeneration studies remain the golden standard of testing
• Advances in animal research are recommended to support the discovery of novel therapies
• Animal skeleton features, the study models, the assessment plan, dosing and statistics should be considered
1. Introduction

The already high incidence of bone trauma in the human population will inevitably increase as the human population ages. Osteoporosis as the major underlying condition makes approximately 27.6 million men and women in the EU (6% of men and 21% of women aged 50–84 years) to be susceptible to a bone fracture [1]. In 2010, approximately 3.5 million bone fractures were reported in the EU with direct healthcare costs of €37 billion and 1.180,000 quality adjusted life years lost [2]; these costs expected to undergo a 25% increase by 2025. Large bone defects as well as non-unions and extensive bone loss after fractures still remain challenges for efficient clinical interventions and require additional support of the damaged site. Because present therapeutic approaches are often accompanied with prolonged treatments, pain and risk of infection, haemorrhage, nerve damage and loss of function, there is a significant unmet medical need for the development of new options for bone repair and the prevention of bone non-unions. Various animal models are available to study the efficacy, safety and tolerability of new therapies.

The objective of this review is to provide an overview of bone defect animal models and available tools for the assessment of bone healing, as well as to suggest guidelines for rational animal use in an attempt to advance bone research as well as to support the development of investigational products in bone regeneration.

2. Bone regenerative strategies: biomolecules, cells and biomaterials

Bone healing is a precisely orchestrated regenerative process, which restores the bone quality by mimicking embryological cascade of events. Bone healing process is traditionally divided into three stages: an early inflammatory stage, a repair stage and late remodelling [3]. A schematic presentation of a long bone healing stages and grades are presented in Figure 1A.

Although bone possesses endogenous self-repair mechanisms [4-8], in conditions such as impaired blood supply, excessive damage to the periosteum, inadequate immobilization, infection at the affected area, mineral and vitamin deficiencies, underlying diseases and side effects of certain medications and radiation, the enhancement of the regenerative processes is necessary to ensure the rapid and adequate restoration of skeletal functions [9-11]. The standard therapy to treat bone fractures/defects includes mechanical support either via cast and/or mechanical devices (e.g. nails, plates and screws). Additional strategies being used and currently developed to further support bone healing are primarily based on the use of: (1) active ingredients (biomolecules), (2) cellular therapies and (3) biomaterials.
Figure 1. A: Tool for assessment and follow up of a long bone healing along the bone healing cascade expressed as a healing grade [12,13] and % of bone healed. The early phase of healing (grade 0-3) enables temporarily fracture stabilization and further endochondral bone formation, and is characterized by the recruitment of mesenchymal cells and successive chondrogenesis resulting in a soft callus formation [14]. The second stage (grade 4-5) is distinguished by deposition of the collagen and subsequent mineralization resulting in a woven bone formation. The last stage (grade 6) is characterized by the bone remodelling which restores the original bone structure and strength. The assessment of the healing process determines the sample size and end-points when planning an experiment. The scheme exemplifies a non-critical size defect healing. If one assumes that it depicts control animals, and that the grades “0” and “1” indicate average scores for an early process and grade “6” indicates an average score of a late process (6-8 weeks for a rat), then grades 0-1 or 6 would be biased towards “no difference” between a control and treatment intervention. Hence, “mid-time” evaluation points (grades 2-5) are of interest for a comparison (see section 5 for more details). B: Suggested time points for the assessment of bone fracture healing in the mouse, rat, rabbit, pig, sheep, dog and non-human primates (NHP). Overall duration of an experiment and assessment time points in critical size defect studies should be extended for at least 30-40% of time used for a bone fracture (see section 5, Figure 3A,B).
2.1. Biomolecules

Biomolecules used in the regenerative therapies for bone are mainly various growth factors [15]. Osteogenic factors primarily belong to the TGF-β superfamily, and the most studied factors are bone morphogenetic protein BMP2, BMP4, BMP6 and BMP7 [16,17]. Because vascularization is essential for bone regeneration, angiogenic factors VEGF, PDGF, FGF and IGF are also being extensively tested for their usefulness in bone repair [18-24]. Immunomodulatory and anti-inflammatory agents, such as selective anti-cytokine therapies, corticosteroids and non-steroidal anti-inflammatory drugs, are used to direct specific effects on the regeneration and resorption pathways during bone healing [25,26]. Additionally, the use of parathyroidal hormone (PTH), growth hormone, steroids, calcitonin and vitamin D in systemic applications has also been shown to advance bone healing by stimulating osteogenesis, angiogenesis and osteoblast differentiation [27-30]. Various combinations of biomolecules have also been extensively evaluated in pre-clinical models with mostly positive results [27,31-37].

2.2. Cell-based therapy

Cell-based therapy utilizes stem/progenitor mesenchymal cells originally identified among bone marrow stromal cells [38]. Although most studies were conducted with bone marrow derived mesenchymal progenitor cells (MPCs), other tissues have been described to contain osteoprogenitor cells with similar regenerative potential including adipose tissue, muscle, umbilical cord blood, periosteum, dental pulp and periodontal ligament [4,39-45]. The multilineage differentiation ability, paracrine effects and immunomodulatory properties of MPCs make them an ideal for tissue engineering and regenerative purposes [5,7,46-48]. Under appropriate conditions MPCs could be differentiated into a variety of mesenchymal tissues such as bone, cartilage, tendon, ligament, marrow stroma, muscle, fat and dermis [4,49-53]. To induce fracture healing, MPCs are expanded ex vivo prior to their autologous grafting to the fracture site and differentiated into osteogenic lineages to promote bone regeneration [9,54]. Such cell-based strategy approaches have been used to demonstrate that autologous bone marrow-derived MPC transplantation was superior compared to unloaded scaffold [5,7]. Unique immunological characteristics of MPCs suggest that the implantation of allogenic or xenogenic MPCs could be successfully used for a cell-based therapy. Cells of non-mesenchymal origin such as endothelial progenitor cells may also enhance bone regeneration by secreting paracrine osteoinductive and angiogenic factors [7]. Additionally, tissue engineering, a process of developing biological tissue substituents for restoring, maintaining or improving tissue function [55], were used to construct a single device combining all of the important components of bone repair (osteoconductive scaffold, osteoinductive growth factors and osteogenic cells) [4,5,7,10,55-57]. Recent studies have attempted to improve the basic protocols of cell-based therapy via additional tissue engineering strategies, including alternative osteoprogenitor population, gene delivery modification of MPCs, growth factors or pharmacological compounds. Currently, more than
20 different clinical trials involving bone tissue engineering approaches using cell therapies are reported in Clinicaltrials.gov (www.clinicaltrials.gov).

2.3. Biomaterials

Starting from the natural materials such as bovine collagen which is currently raising safety concerns [58], the field of orthopaedic biomaterials has expanded to include an impressive array of materials that are currently being tested in preclinical models [59]. Biomaterials have a range of properties, from osteoinductive and osteoconductive to immunomodulatory. Hydroxyapatite and calcium phosphate as well as their composites such as HA/poly(DL-lactic-co-glycolic acid) (PLGA), in the form of ceramics, cements and coatings have shown osteoinduction in animal models [60-66]. Various hybrid materials combined as co-polymers, polymer blends and polymer-ceramic blends have also shown efficacy [67-73]. Advanced hydrogels, naturally derived collagen and gelatin gels as well as synthetic polyethylene glycol and poly–vinyl alcohol-based hydrogels, serve as matrices for other products and mimic the extracellular matrix topography [74-76]. Biomaterials with immunomodulatory strategies, such as artificial extracellular matrices (ECMs) (hydrogels, ECM coatings) and materials with surface property modulation, have the ability to modify the immune function and improve bone repair and regeneration [25,77].

3. Overview of Methodology and Animal Models for Bone-healing Studies

Working with animals is a privilege and scientist as well as their institutional ethical boards should do their best to conform to the current animal care guidelines [78,79]. In recent years, the regulatory and scientific community imposed stringent rules to ensure that the wellbeing of laboratory animals is respected and that the 3R’s paradigm implemented whenever possible without compromising the quality of the study and data analyses [80,81]. Despite copious improvements, too many animal studies are still being published without regard of their poor design, use of inadequate or insufficient biomarkers, inaccurate conclusions and/or producing insignificant data. The overall benefit of these studies to scientific community is negligible, with published studies being redundant and offering very little novelty. For instance, the BMP preclinical development is a telling example of the misuse of laboratory animals due to non-existent standard operating procedures for testing of novel compounds for bone regeneration. Large numbers of animal models, species and doses have been used to support the development of BMP2 and BMP7 devices, however majority of these studies yielded inconclusive results that were misleading and difficult to interpret. Between 1988 and 2004 hundreds of experiments were conducted on more than 17.000 animals (literature search was performed via PubMed using terms bone morphogenetic protein 7 and bone morphogenetic protein 2, revealing 157 and 421 articles, respectively). In our opinion, literature based data represent only a small fraction of the total number of animals used. Despite this, after years of use in clinics both BMP devices have
been confronted with major side-effects and their clinical use has been recently scrutinized [17,82-84].

3.1. **Principles of Study Design**

The analysis of hard tissues requires a long, complex and expensive experimentation. Scientists, clinicians and regulatory authorities have recognized the necessity of using two laboratory species and several independent biomarkers when assessing the effect of novel bone and fracture-healing therapies [85,86]. The first step when preparing a study is to precisely determine the study goals and establish the criteria used to evaluate the overall success of the study. The following step is to plan the study design in which ten essentials should be selected: 1) Animal model (optimizing goals of the study); 2) Animal species; 3) Animal sex and age; 4) Study duration (allowing for the biological process to initiate and complete); 5) Number of animals per study group (sufficient for statistical analyses); 6) Dose and route of administration of the test article (mimicking anticipated clinical use); 7) Appropriate controls (including a sham, vehicle and/or a “positive” control group to ensure the credibility and reproducibility of the data and a standard of care drug with a well-known efficacy/safety should be used as a “positive” control); 8) Supply of the test article (sufficient for the entire study); 9) Optimal *in vivo* and *ex vivo* biomarkers, and 10) Tissue collection, storage and analyses planning. The common wisdom of *in vivo* experimentation is often ignored for various reasons, most frequently due to a lack of experience and poor planning, a lack of funds, a lack of in-house expertise, short timelines and the inadequate selection of biomarkers. The publication of poorly designed studies on animals should be restricted for both ethical and scientific reasons.

3.2. **Methods and Technology**

All currently marketed drugs for the treatment of skeletal disorders, including fractures were successfully tested in preclinical models. The value of preclinical work involving animal models depends on two essentials. The first determinant depends on the availability of an animal model that mimics a human disease involving bone repair so that the data generated can be used to predict the drug efficacy and safety in patients. Examples of animal models with good predictability of clinical outcomes include models of postmenopausal osteoporosis [87-92], models of glucocorticoid induced bone loss [93-96], models of cancer metastasis to bone [97], disuse models [98,99], fracture healing models [100-103] and several others. The second determinant of successful experimentation relates to translational biomarkers of novel therapy efficacy and safety that can be accurately predicted and monitored in patients. Numerous methods that are thoroughly understood, extensively described and tested for predictability are available for testing the efficacy and safety of novel treatment targets (Table 1).
Table 1. The “toolbox” of methods and technologies that is available for in vivo and ex vivo assessment of bone physiology and pathology.

**IN VIVO ASSESSMENT**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Process/Parameter</th>
<th>Assay/Technology</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomarkers in serum and urine</td>
<td>Bone formation, resorption, metabolism, cartilage formation, connective tissue degradation, calcitropic hormones</td>
<td>P1NP, Osteocalcin, BSAP, CTX, TRAP5b, Ca(^{2+}), P(^{2-}), PIIANP, ICTP, IGF-1, PTH, Vit. D, calcitonin, T3/T4</td>
<td>High</td>
</tr>
<tr>
<td>Imaging technologies</td>
<td>Bone anatomy, bone mass, geometry, and structure</td>
<td>Standard radiology, DEXA, pQCT, micro-CT, MRI, PET, use of contrast agents</td>
<td>High</td>
</tr>
<tr>
<td>Functional tests</td>
<td>Biomechanic/Biometric</td>
<td>Dynamic Weight Bearing System, Gait analyses (Digigait), some other</td>
<td>High</td>
</tr>
<tr>
<td>Mechanical properties</td>
<td>Bone strength: maximum load, stiffness, toughness, ultimate strength</td>
<td>BioDent</td>
<td>Too early</td>
</tr>
</tbody>
</table>

**EX VIVO ASSESSMENT**

<table>
<thead>
<tr>
<th>Test</th>
<th>Activity/Parameter</th>
<th>Assay/Technology</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaging technologies</td>
<td>Local bone anatomy, bone geometry, bone mass, bone structure</td>
<td>Standard radiology, DEXA, pQCT, micro-CT, MRI, PET</td>
<td>High</td>
</tr>
<tr>
<td>Bone Biomechanics (strength)</td>
<td>Bone geometry, composition and strength of cortical and/or cancellous bone</td>
<td>Various methods (3- and 4-point bending methods, tensional test, compression test, Finite element modeling)</td>
<td>High</td>
</tr>
<tr>
<td>Undecalcified bone histology and histomorphometry</td>
<td>Bone remodeling and modeling at cortical bone envelopes, cancellous bone, Bone Formation Rate, Mineral Appositional Rate, osteoid, mineral</td>
<td>Requires in vivo labeling with fluorescent markers, embedding in methylmetacrylate, cutting and analyses (histomorphometry) or staining (von Kossa, Goldner trichrome, Toluidine blue). Cryosections should be considered.</td>
<td>High (if bone biopsy is available)</td>
</tr>
<tr>
<td>Decalcified bone histology and histochemistry</td>
<td>Bone cells (osteoblasts, osteoclasts, bone lining cells, osteocytes, bone marrow, chondrocytes</td>
<td>Routine bone and joint stains (H&amp;E, Toluidne blue, Safarin O) or immune-staining (TRAP, Factor VIII, PCNA, PGP9.5)</td>
<td>High (if bone biopsy is available)</td>
</tr>
<tr>
<td>Microscopy</td>
<td>Bone structure, lamellar bone, woven bone, cellular analyses</td>
<td>Polarized microscope, electron microscope</td>
<td>High</td>
</tr>
</tbody>
</table>

P1NP - serum type 1 procollagen (C-terminal/N-terminal); BSAP - serum bone-specific alkaline phosphatase; CTX - collagen type 1 cross-linked C-telopeptide; TRAP5b - tartrate-resistant acid phosphatase 5b; PIIANP - type I procollagen N-terminal propeptide; ICTP - carboxyterminal telopeptide of type I collagen; IGF-1 - insulin-like growth factor-1; PTH - parathyroid hormone; T3/T4 - thyroxine/thyroid hormone; PCNA - proliferating cell nuclear antigen; PGP9.5 - neuron cytoplasmic protein gene product also known as ubiquitin C-terminal hydrolase 1 (UCHL-1).
Radiological methods based on detecting bone minerals should always be used in combination with histology. Quantitative computed tomography (qCT), peripheral quantitative CT (pQCT) and micro CT (μCT) are additional techniques that are superior to X-ray in assessing bone geometry, mass and structure. X-ray with/without pQCT is the method of choice for in vivo follow-up due to its good correlation with the mechanical testing of bone strength [104]. The strength of bone and callus assessed via the 3- and/or 4-point bending method (or torsional testing performed ex vivo) supplements the assessment of bone regeneration quality and provides opportunities to correlate and accurately interpret radiologic and other data (serum biomarkers, histology). Bone histology and histochemistry performed ex vivo provide key and quantifiable methods to test bone cell activity during bone regeneration and should be an integral part of bone healing animal studies. Undecalcified bone histomorphometry based on the use of fluorescent dyes such as calcein, tetracycline and alizarin labelling mineralizing bone surfaces provides an accurate quantification of the bone formation and resorption processes in the callus and in the surrounding cortical and cancellous bone. Differential staining using the van Kossa method for minerals and counterstaining with toluidine blue enables the meticulous evaluation of cartilage and mineralized bone at the callus. The same metacrylate embedded bone samples can also be used for polarized light microscopy for determining the lamellar or woven bone structure in the newly formed bone. Examples of the healing process in long bones by using radiology methods and histology is depicted in Figure 1B. Both methodologies can be used separately or in combination to accurately score the healing cascade in order to assess efficacy and safety of tested therapies.

Serum and urine biomarkers of bone formation and resorption are highly desirable and recommended for the assessment of bone metabolism. Although serum biomarkers can be used to measure the overall activity of bone cells throughout the skeleton, and may not always detect local changes in bone activity around the fracture, these markers have a great translational value and as such provide an exceptional tool for accurately monitoring skeletal metabolism in live animals.

4. **3R: Replacing, reducing and refining to improve animal welfare**

Although scientists are continually developing more complex alternative techniques, like engineered organs and in vitro tissue models, animal models are still the gold standard for fracture healing testing due to their ability to mimic complex human physiological processes and bone mechanics which cannot be simulated and replaced by even most advanced non-animal technologies. This is particularly true for testing of new medicinal products and during preclinical phases of drug development. Contemporary biomedical research projects are increasingly encountering efforts to apply the principles of the “3Rs” for conducting humane experimentation in animals focused on Replacement, Reduction and Refinement [80,81].
To align more firmly with the principles of the 3R, EU member states adopted the 2010/63/EU directive and the same processes are being implemented by other policy makers: EMA (http://www.ema.europa.eu/ema/), the OECD (http://www.oecd.org/health/) and the ICH (http://www.ich.org/products/guidelines.html/).

4.1. Animal model considerations

Animal models of bone regeneration fall into two categories: (1) ectopic models are primarily used to distinguish between the proliferative and inductive capacity of new products, while (2) orthotopic models are used to test the efficacy and safety of the new products and/or procedures. Ectopic models are relatively simple, less costly and less invasive. For example, ectopic bone formation models were recently used to elucidate whether the maturation status of implanted cells determines the origin of tissue-engineered bone [105] and to demonstrate new bone formation after the implantation of the OSTEOGROW device (rhBMP6 in autologous blood coagulum) without any signs of inflammation or fibrosis [17]. Although successfully used as a preliminary model for screening various formulations of osteogenic cells, scaffolds and growth factors, the ectopic model displays serious limitations including the eventual reabsorption of newly formed bone and the lack of effective mechanical stimulus required for bone remodelling [106]. Orthotopic models represent investigational procedures performed in or around the bone itself. The classification of the bone healing models is presented in Figure 2. The process of bone repair can also be studied in models of bone disuse under various loading conditions via animal models in which one limb is fully or only partially deprived of weight bearing activity. These models are more sophisticated because they require multiple procedures such as sciadic neurectomy, amputation or various immobilization techniques combined with bone defects [107,108]. The choice of the model should be based on scientific, ethical and practical merits of a particular study and reflect the human biology or disease and should accommodate for appropriate clinical settings with relevance to the product being tested [101,109-111].
Figure 2. Graphical presentation of animal bone models based on the investigated effect of the therapy. Ectopic models based on injection site are referenced as subcutaneous (s.c.), intramuscular (i.m.) and under the kidney capsule implantation (u.k.c.) models [16,112]. Based on the creation of the defect size, orthotopic models of bone healing are divided in a non-critical size defect models with the capacity to heal without an intervention, whereas in the critical size defect model, bone regeneration and restoration of the function will not occur without an intervention. Orthotopic models can also be grouped based on the anatomical location in the body (appendicular and axial skeleton) but also based on the bone structure and microanatomy of the bone being studied (cortical and/or cancellous bone). When based on mechanical loads imposed on the studied bone, animal models can resemble physiological and non-physiological conditions. For example, bone repair can be studied in non-loaded bones (such as calvarias), in bones that experience normal weight bearing loads (appendicular (vertebrae) or axial (long bones) skeleton) or in the mandible where non-weight bearing intermittent loads occur during chewing [113,114]. Bone defect studies can be performed in traditional laboratory animals (mice, rats, rabbits, dogs and non-human primates - NHP) as well as in domestic animals such as sheep and pigs, which are often used as viable substitutes for dogs and NHP.
Systematic factors discussed in the section 3.1 of this manuscript should be taken into account during the planning phase of the study. Models that are better understood and described in the literature often have proven records of being more predictive of clinical outcomes. The age of study animals as well as gender can influence the bone repair through the action of calcitropic hormones and thus merit careful consideration [115-119]. For example, aged, thyroparathyroidectomized and ovariectomized (OVX) animals are known for delayed fracture healing and reduced bone mineral density; therefore, OVX animals are frequently utilized to study osteoporotic fractures because these models mimic postmenopausal women [88,120,121]. Historic evidence suggests that rats, rabbits and mice are the most frequently used species to study bone physiology and drug efficacy and safety accounting for approximately 80% of all animals used to study bone repair, while other species including sheep, goats, pigs, dogs and non-human primates (NHP) make up for the remaining 20% [122,123]. The choice of a particular species is a critical step and is often based on the biochemical and microstructural characteristics of the bone tissue as well as on the similarities of the healing processes between the particular species and humans (Table 2, Figure 1). In addition, difference in gross anatomy as well as differences in distribution of the cortical and cancellous bone compartment within each bone should be taken into account when deciding which species to choose for the particular study since those differences reflect biomechanical properties that play a critical role in the bone repair processes [124]. The availability of serum biomarkers and the translatability of biomarkers to the clinical environment is a very important issue. Rats, dogs and NHP are routinely used in preclinical safety studies and serum biomarkers of bone metabolism are well established and validated, therefore using those species is advantageous from a biomarker standpoint; however, the use of dogs and NHP is restricted and those two species should only be used if necessary only for the late stage testing of efficacy and safety parameters. Clearly, no animal model entirely mimics human conditions because no animal species has a skeletal or biomechanical properties identical to human. Additional factors that could influence selection of the animal model include the size of the animal, the cost to acquire and care for animals, animal availability, ethical acceptability, tolerance to captivity, breeding cycles, ease of housing and handling, adequate facilities and qualified staff and familiarity with the model, technical capacities etc. [109,110,125].
Table 2. Bone research related characteristics of laboratory animals.

<table>
<thead>
<tr>
<th>Laboratory animal</th>
<th>Life expectancy (months)</th>
<th>Bone maturation (epiphyseal plate closure) (months)</th>
<th>Time to union of fractures (weeks)</th>
<th>Similarity to human bone (Ma+Mi+C+R)</th>
<th>Ethical acceptance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>18-36</td>
<td>5</td>
<td>3</td>
<td>0+1+0+1 = 2</td>
<td>high</td>
</tr>
<tr>
<td>Rat (Sprague-Dawley)</td>
<td>30-48</td>
<td>11</td>
<td>4-6</td>
<td>1+1+0+1 = 3</td>
<td>high</td>
</tr>
<tr>
<td>Rabbit (New Zealand White)</td>
<td>84-96</td>
<td>6,8 tibia, 5,3 femur</td>
<td>6-7</td>
<td>1+1+2+1 = 5</td>
<td>high</td>
</tr>
<tr>
<td>Dog (Grayhound)</td>
<td>108-168</td>
<td>7,5 tibia, 7,3 femur</td>
<td>10-13</td>
<td>2+2+3+2 = 9</td>
<td>low</td>
</tr>
<tr>
<td>Sheep (Suffolk x Dorset)</td>
<td>180</td>
<td>17</td>
<td>10-14</td>
<td>3+1+2+2 = 8</td>
<td>medium</td>
</tr>
<tr>
<td>Pig (Gottingen)</td>
<td>180+</td>
<td>12-24, 28 femur</td>
<td>12-24</td>
<td>2+2+3+3 = 10</td>
<td>medium</td>
</tr>
<tr>
<td>Monkey (Rhesus)</td>
<td>240-360</td>
<td>75 femur</td>
<td>16-24</td>
<td>3+3+3+3 = 12</td>
<td>low</td>
</tr>
<tr>
<td>Human</td>
<td>47-83 years¹</td>
<td>240</td>
<td>18-24</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

¹ WHO Life expectancy at birth, both sexes, 2011; ² ref [126]; ³ ref [127]; ⁴ ref [128]; ⁵ ref [129]; ⁶ adapted from [110] – comparison based on macrostructure (Ma), microstructure (Mi), composition (C) and remodelling (R) and scoring from 0 (not similar)-3 (very similar) resulting in the total sum representing similarity.

4.2. Species Skeleton Specification

4.2.1. Rodents

Of all laboratory animals mice and rats are considered to have skeletons and bone biologies that are least similar to humans since their skeletons are modelling-driven due to permanently open growth plates at the epiphyses of long bones, a lack of the Haversian system and low cancellous bone content at the epiphyses of the long bones [100,130-136]. However, studies in rodents are very informative and cost effective. Rodents, primarily mice are genetically very well defined and are best suited for studies with genetically modified strains to address specific molecular mechanisms of bone physiology and pathology. The advantages of using rodents, and particularly rats in early stage studies are numerous and include broad availability, inexpensive housing, easy handling, small size (relatively small quantities of test article needed), well-defined and described procedures and models and biomarker availability. Rats are most regularly used in safety toxicology studies and also to study the pharmacodynamic and pharmacokinetic properties of novel treatments [137]. Despite some deficiencies, rat models of skeletal diseases are very predictive of drug efficacy and safety in humans because bone cells, osteoblasts, osteocytes and osteoclasts in rats have similar receptors to human bone cells and therefore react to drug challenge in a similar fashion to human bone [85]. Some characteristics of rodent models, such as an open growth
Plate, may be unfavourable for studies focusing on the adult skeleton, however, studies in rodents are very useful when investigating efficacy and safety of drugs targeting juveniles. ICH-harmonized guidelines for biotechnology-derived pharmaceuticals were recently updated to stress the importance of selecting the relevant species for the investigational product biological activity as well as for the safety [138].

4.2.2. Rabbit

Rabbits are the smallest commonly used non-rodent species in musculoskeletal research studies [139] because the rabbit skeleton does not include the two major drawbacks of rodent models; the lack of the Haversian system and permanently open growth plates. Rabbit bone also differs from that of humans in its size and shape [124]. Histologically, the skeleton of rabbits consists of primary lamellar bone while vascular canals parallel the long bone axis. Rabbits are commonly used models for bone healing studies due to the vast experience with rabbit handling, short duration necessary to reach the mature bone characteristics and bone densities that are similar to those of humans [103,140]. Rabbit models were used to study metaphyseal fractures [141], mandibular distraction osteogenesis [142], mandibular defect repair [143], investigational products in cranial models [144], critical size defects in animals of different age [145] and spinal fusion [146,147]. Rabbits are also widely used to study novel biomaterials, growth factors and stem-cells approaches [142-145,148,149].

4.2.3. Dog

Canine bone healing models are frequently used in musculoskeletal and dental research because significant amounts of information are available regarding the predictability of these models for human conditions [110,140,150]. Dog bones have a mixed microstructure with a primarily secondary osteonal bone and a plexiform bone in the vicinity of the periosteum and endosteum [140]. The bone composition in dogs is similar to that of humans and the structural properties of bones from several skeletal sites was recently summarized by Bagi et al. [124]. Dogs are often used in safety studies as a second species and biomarkers of bone metabolism are validated and are well-established for dogs. However, there are societal concerns regarding the use of dogs in biomedical research and the ethical framework related to their welfare [151] consistently restricts the use of dogs in preclinical testing.

4.2.4. Pig

Although pigs have been used for decades in bone studies this model was never widely deployed due to the fact that commercial breeds usually grow quickly and reach extreme body weights. With advances in the breeding of minipigs and micropigs, the use of pigs in biomedical and orthopaedic research has increased [152]. Regarding bone anatomy, microstructure, remodelling and healing, porcine bone closely resembles human bone [153,154]. Pigs were found to exhibit spontaneous vertebral fracture and their rates of bone
removal and deposition (trabecular and cortical bones) are similar to humans although porcine bone remodels slightly faster than human bone [152,155,156]. Limitations for the use of NHP in regulatory toxicology studies have opened discussions on the suitability of using minipigs in drug development studies [157]. Recently, it was shown that minipigs are comparable to NHP in immunogenicity testing [158,159] and their liver metabolism is similar to that of humans [160]. Additionally, models of porcine osteoporosis have been developed to expand the usefulness of this species for bone research [161-164].

4.2.5. Sheep

Sheep tibia models are considered to be valid and reliable for the evaluation of bone regeneration, with the advantage of having a maximal weight bearing scenario and long bone dimensions in adult animals that are suitable for the testing of human implants and prostheses [165-168]. However, sheep are seasonal breeders so their bone metabolism changes during the year which presents a significant hurdle for bone metabolism studies. The bone maturation period in sheep is long and the microstructure in the young animals is distinct (plexiform bone). In adult sheep, the bone structure is different from humans, consists primarily of primary bone and Haversian remodelling occurs during adulthood [110]. The bone mineral density and bone strength in sheep is increased relative to human. Moreover, sheep models of critical size defects are extensively applied to evaluate cell-based therapeutic approaches using autologues, allogeneic or xenogeneic MPCs in combination with growth factors and different types of scaffolds to enhance bone regeneration showing significant advantages compared with cell-unloaded (empty) scaffolds [43,53,166,169-176].

4.2.6. Non-Human Primates

NHP are the best characterized large animal model for skeletal research and their skeleton and posture as well as their bone structure, composition and remodelling patterns is similar to those of humans [124,177-179]. In addition to a high similarity between monkeys and humans regarding drug metabolism between monkey and human, the existence of validated serum and urine biomarkers with high translational value to human is of the outmost importance for regulatory studies using NHP. Although skeletal studies in Old World primates yield valuable data, the use of NHP are constrained by ethical and technical considerations, including high cost, limited availability and regulations (Directive 2010/63/EU) [80,180]. Although very useful in bone research for novel therapies, NHPs should only be used in situations when efficacy, safety and toxicity studies in other species could not provide appropriate answers i.e. human antibodies or indications such as hereditary non-union of the tibia or long bone fibrodisplasia. Marmosets were recently proposed as good alternative for skeleton studies because the adult marmoset skeleton has similar anatomical characteristics that are similar to those of adult humans, including the absence of growth plates, the presence of Haversian system and true remodeling of cancellous and cortical bone [181,182]. Compared with macaques, marmoset monkeys have an earlier puberty and sexual maturity and presumably achieve earlier peak bone mass. They
are easy to breed and to handle under controlled laboratory housing conditions. Similar to FDA guidelines [183], in the EMA Guideline on the evaluation of new medical products in the treatment of primary osteoporosis [184] states, that these substances should be tested in at least two species, one of which should be an ovariectomized rat and the other an animal with evaluable cortical bone remodeling. Primates, sheep and pigs are suggested as a second animal model by the EMA. Common marmoset monkeys (*Callithrix jacchus*) also fulfill these requirements because they show osteonal remodeling that is very similar to that of humans. It is therefore highly recommended that institutions involved in drug development request a scientific advice from regulatory agencies for opinions regarding non-clinical data requirements prior to firsts-in-human studies [184].

4.3. Animal models in cell-based therapies

To be able to utilize the major advantages of experimentation in laboratory animals, models of bone repair need to be optimized in accordance to the unique characteristics and requirements of different species. Different cell-based therapy approaches to treat segmental defects in weight-bearing long bones are given in Table 3 to illustrate the great variability of models regarding cell population and scaffold selection, protocols of intra-defect transplantation, species and defect localization, follow-up period and outcome measurements. Considering the wide diversity of conducted research, investigators should be extremely cautious to translate conclusion in-between models and species and pay special attention in designing the animal studies to draw valuable and reproducible results.
Table 3. Different cell-based therapeutic approaches used for the treatment of critical size segmental long bone defect in small and large animal models

<table>
<thead>
<tr>
<th>Cell-based therapy</th>
<th>Species (study reference and details)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rodent</td>
</tr>
<tr>
<td>autologous (syngeneic) or allogeneic BM-derived MPCs</td>
<td>Mouse: [185], [186]</td>
</tr>
<tr>
<td></td>
<td>Rat: [187] (DM), [188]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>human BM-derived MPCs</td>
<td>Mouse: [198]</td>
</tr>
<tr>
<td></td>
<td>Rat: [199], [200], [201]</td>
</tr>
<tr>
<td>non BM-derived MPCs or non-MPC cell-based therapy(^1)</td>
<td>Mouse: [204] (hUCPVC), [205] (hSDF1/hBMP2-mFTG)</td>
</tr>
<tr>
<td></td>
<td>Rat: [206] (rEPC), [207] (hADPVC)</td>
</tr>
<tr>
<td>genetically modified or labeled MPCs(^2)</td>
<td>Mouse: [211], [212] (hBMP2-mMPC), [213] (hMPC/GFPCol1α1), [214] (mMPC/GFP), [215] (ShhN-MPDMPC)</td>
</tr>
<tr>
<td></td>
<td>Rat: [216] (BMP7-rDF), [217] (hBMP2-rMPC), [45] (hBMP2-MDC or hBMP2-ADC)</td>
</tr>
<tr>
<td>MPCs with growth factors/compounds</td>
<td>Mouse: [223] (hVEGF/hMPC), [224] (hVEGF/hBMP2/hMPC)</td>
</tr>
<tr>
<td></td>
<td>Rat: [225] (hBMP2/hMPC), [226] (hBMP7/hMPC)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Cells were genetically modified by viral or non-viral transfection to overexpress different growth factors or regulatory molecules. Cells transfected with US2/US3 genes downregulated MHC I expression. In some studies MPCs were fluorescently labeled for in vivo tracking.

\(^2\) Non bone marrow-derived MPCs were isolated form adipose tissue, periosteum, muscle, articular cartilage, placenta, human umbilical cord blood, etc. Therapy with cells other than MPCs used endothelial progenitor cells or genetically modified fat cells.

Abbreviations: BM - bone marrow; h - human; m - mouse; r - rat; rb - rabbit; s - sheep; MPCs - mesenchymal progenitor cells; GFP - green fluorescent protein; DM - diabetes mellitus prone strain; BMP - bone morphogenetic protein; VEGF - vascular endothelial growth factor; PRP - platelet-rich plasma; Shhn - N-terminal sonic hedgehog peptide; UCPVC - umbilical cord perivascular cells; EPC - endothelial progenitor cells; bFGF - basic fibroblast growth factor; Ang1 - angiopoietin 1; MHC - major histocompatibility complex; SDF1 - stromal cell derived factor 1; FT - fat tissue graft; DFAT - adipocyte-derived dedifferentiated fat; Col1α1 - Collagen, type I, alpha 1; PDMPC - periosteal-derived mesenchymal progenitor cells; MDC - muscle-derived cells; ADC - adipose tissue-derived cells; DF - dermal fibroblasts; ADPVC - adipose tissue-derived perivascular cells, PMPC - placenta-derived mesenchymal progenitor cells

5. Experimental design and statistical considerations for animal fracture models

Here we describe a few possible modes for the standardization of the non-clinical efficacy evaluation of new treatments using the examples of critical and non-critical size defects of long bones in rat and rabbit models. We suggest that in both paradigms \(\mu\)CT (in vivo or ex-vivo; bone volume/tissue volume ratio), X-ray (rabbit in vivo) and histology are used; the latter two employing the elaborated scoring system shown in Figure 1. These methods and measures are reliable and reproducible with comparable mild-to-moderate variability in treated and control animals with relative standard deviation (RSD) in the range of 10-30% [231,232].
5.1. Critical size defect

Because the defect does not heal spontaneously and remains practically unchanged in control animals, evaluation at any time-point after the initial 1-2 weeks could indicate a bone-healing effect. However, we suggest that the evaluation of new treatments should be compared with the approved treatments, with main assessments based on repeated in vivo radiological measures over a period extending beyond the expected time of physiological union seconded in non-critical size fractures focusing on the overall process (Figure 1). Experiments should include only a few previously demonstrated inactive control animals (“placebo” to the new treatment). Figure 3A depicts four repeated radiograms and μCT scans (table in Figure 3A) taken over time in one such experiment. Data demonstrate low variability of values across the time point. We suggest data analysis by fitting general linear mixed models (GLMMs) with restricted maximum likelihood (REML) estimation (relaxed assumptions as compared to repeated measures ANOVA, units with individual missing values not excluded) to produce time-averaged difference between treatments. An alternative analysis would be based on integration of the response as the area under the curve (AUC) of change of the radiological measure (a difference between newly formed bone at each time point vs. 0) by fitting a general linear model (GLM). We suggest that the defect size is considered as a covariate. Figure 3A, B depicts a new treatment superior to an approved one (bone formation 4.6 fold greater based on time-averaged response and 2.9 fold greater based on AUC). We suggest that a new treatment indicates a potential superiority if any of these measures is at least 25% in its favour. With four repeated assessments, assuming variability as depicted in Figure 3 (or relative standard deviation (RSD) around 20% at each time point, as well as for AUC) autocorrelation 0.6 and autoregressive (1) [AR(1)] covariance structure, 7 animals per group per time-point provide 80% power to detect such a difference at a two-sided $\alpha=0.05$. The same sample size applies for 25% difference in AUC. Adding a treatment time interaction term in GLMM allows for comparisons at different time points, but requires adjustments of comparison-wise alpha level, resulting in a need for a larger sample (e.g. for pairwise comparison at four time points, under the above conditions 17-18 animals per group). To detect a time-averaged or AUC difference of at least 15% under the above conditions, 17-20 animals per group are needed. If such an experiment fails to reject null-hypothesis, it is reasonable to conclude that the test is comparable to the reference. We suggest that if the entire 95% confidence interval around the difference falls within the range -15% to +15%, it is plausible to conclude equivalence of two treatments. With 17-20 animals per group, the experiment provides >80% power to even formally demonstrate equivalence under the above conditions.

5.2. Non-critical size defect

The same approach with repeated assessment in vivo can be applied to the non-critical size defect model. We suggest that animals should be scanned immediately after the fracture – if non-zero values are present and an adjustment for baseline should be considered. Since the
defect heals spontaneously, assessment very early or late in the process are biased towards “no difference”. It could also be difficult to distinguish between different active treatments. We suggest that a test treatment (T) should be considered effective if it yields by at least 30% higher healing “scores” around mid-time of the spontaneous process as compared to an inactive control (Ctr). Figure 3C depicts an experiment in which several active and “disruptive” treatments were assessed along with the inactive control at the single time-point late in the healing process (ex vivo rat bone, μCT). Data analysis (GLM) requires adjustment for pairwise comparisons. Figure 3D depicts a hypothetical experiment in which a treatment is evaluated against an inactive control ex vivo at two time points around the mid-time of the process. If comprises one between-group factor with four levels (2 treatments x 2 time points) and should be analysed in a GLM with adjustments for two post-hoc comparisons of interest (T-1 vs. Ctr-1; T-2 vs. Ctr-2). With group RSD of 20%, 10 animals per group are needed (40 total) to detect a 30 % difference between T and Ctr at any of the two assessments.
Figure 3. A. X-ray of critical size defects of rabbit ulnae treated with a BMP6 in the modified whole blood coagulum (WBCD) and commercial BMP7 device and assessed at 0, 2, 4, 6 and 8 weeks after surgery. Tabulated data show corresponding BV/TV ratios obtained by μCT [233]. B. Graphical representation of BV/TV% over time (mean±SD; n=8). A GLMM fitted to ln-transformed data to determine time-averaged difference (weeks 2-8) indicated around 4.6 times greater response with BMP6. A GLMM fitted to ln-transformed data to determine the time-averaged difference (weeks 2-8) indicate around 4.6 times greater response with BMP6. A GLM fitted to ln-transformed AUC of BV/TV% difference at weeks 2-8 vs. time 0 indicated around 2.9 times greater response (tabulated GMR). The insert shows a quadratic fit with 95% CI of prediction indicating that the effects of two treatments were not likely to “meet” even after 10+ weeks post-surgery. The intrapolated grey line indicates a hypothetical treatment not relevantly different vs. the approved one (<15% difference). C. μCT imaging of rat femurs ex vivo at 6 weeks after osteotomy in the proximal third of the femur. The rats were randomly assigned to one of the following groups: control; BMP1-3 antibody i.v. (50 ug, 1xwk); BMP1-3 i.v. (3 ug, 3x7wk); BMP6 i.v. (250 ug/kg, 2x/wk) and BMP1-3 (3 ug, 3x/wk) +BMP6 i.v. (250 μg, 2x/wk) therapy (mean±SD, n=6) [234]. D. Outline of a hypothetical experiment in which tested treatment (T) and a control (Ctr) are assessed ex-vivo at two time points (1, 2) around midtime of the spontaneous healing process. Data are independent, i.e., there are four groups of animals (2 treatments x 2 time points). GLMM - general linear mixed models; GMR - geometric mean ratios; AUC - area under the curve; BV - bone volume; TV - tissue volume; GLM - general linear model.
6. **Suggested guidelines for animal use in bone repair experimentation**

To enable the collection of reliable data regarding the efficacy and safety of tested substances, animal models and methods should be carefully selected and combined to guide clinical studies so that the much-needed treatments aimed to facilitate the tissue regeneration process will ultimately reach the patient. Below are suggested guiding principles for conducting animal studies in bone repair scenarios.

1. The skeletal characteristics of each species must be considered when judging the translation of preclinical data to the human clinical situation. The rat model, despite its’ limitations, remains the most informative small animal model and should be the first choice to initiate *in vivo* assessment. If studies in large animals are planned, the best option is to combine efficacy and safety study in the same species. Although NHPs are the best choice due to the high similarity to human outcomes, dogs or pigs are valid alternatives. To bridge the gap between rats and large animals, the use of “intermediary” models (marmoset or rabbit) could be considered with a caveat that these models may not provide sufficient new information to better guide studies in large animals.

2. The dosing and sample size must be based on previous *in vitro* studies as well as the healing biology of a specified indication and healing time point analyses, respectively. 3R principles should always be used in designing experimental protocols.

3. Efficacy testing in animal models should mimic the intended clinical indication.

4. Appropriate methods and markers must be chosen for the assessment of bone healing process as suggested in Table 1. Radiologic techniques in combination with histology, bone mechanics and serum and urine biomarkers are recommended.

5. The scanning of animals prior to recruitment will enable each animal to serve as its own control and ensure screening for potential fractures or malformations, establishing growth plate status and skeletal maturity to avoid biological variation errors.

6. A bone healing assessment plan should be carefully designed while bearing in mind the callus formation time course and the points of biomechanical bone restoration in different animal species. Primary and secondary end points of the study and scoring system should be carefully considered and specified in advance.

7. The chosen model, surgical and therapeutical procedures, variability, sample size and percent of expected outcomes should be supported by detailed and justified pre-experimental statistical analysis.

8. The safety and tolerability of the therapy should be monitored during the study and synchronized with regulatory documents whenever possible.

9. Isolated MPCs should be characterized by phenotype, gene expression profile and functional testing prior the grafting procedure. Markers for labelling the target osteoprogenitor population and lineage tracing approach are particularly useful for the *in vivo* tracking of the transplant. Cell-based engineered constructs could be harvested
at different time-points post-surgery and analyzed ex vivo to evaluate the viability, proliferation and differentiation of the transplanted cells.

10. Cell based therapies in combination with biomolecules and biomaterials require additional evaluation particularly regarding osteogenic differentiation of transplanted cells and graft-host integration. Moreover, construct vascularization, scaffold biodegradation and transplant-host integration are important predictors of skeletal tissue regeneration. Local and systemic immune reactions should also be monitored particularly in the case of allogeneic or xenogeneic grafts.

Acknowledgements

The research leading to some of these results has received funding from the European Community’s Seventh Framework Programme [FP7/2007-2013] under grant agreement n°HEALTH-F4-2011-279239.
Reference List


progenitor cells for the reconstruction of critical sized segmental tibial bone defects in aged sheep, Acta Biomater. 9 (2013) 7874-7884.


