Impact of guidance documents on translational large-animal studies of cartilage repair

Christian G. Pfeifer[1,2,3], **Matthew B. Fisher**[1,2,4], **James L. Carey**[1,2], and **Robert L. Mauck**[1,2,3,6,*]

[1]McKay Orthopaedic Research Laboratory, Department of Orthopaedic Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA [2]Translational Musculoskeletal Research Center, Philadelphia VA Medical Center, Philadelphia, PA 19104, USA [3]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA [4]Department of Biomedical Engineering, University of North Carolina-Chapel Hill, Chapel Hill, NC 27599 and North Carolina State University, Raleigh, NC, 27695, USA [5]Department of Trauma Surgery, Regensburg University Medical Center, Regensburg, Germany [6]Collaborative Research Partner (CRP), Acute Cartilage Injury (ACI) Program of the AO Foundation, Davos, Switzerland

Abstract

Promising therapies for cartilage repair are translated through large animal models toward human application. To guide this work, regulatory agencies publish recommendations ("guidance documents") to direct pivotal large animal studies. These are meant to aid in study design, outline metrics for judging efficacy, and facilitate comparisons between studies. To determine the penetrance of these documents in the field, we synthesized the recommendations of the American Society for Testing and Materials, U.S. Food and Drug Administration, and European Medicines Agency into a scoring system and performed a systematic review of the past 20 years of preclinical cartilage repair studies. Our hypothesis was that the guidance documents would have a significant impact on how large animal cartilage repair studies were performed. A total of 114 publications meeting our inclusion criteria were reviewed for adherence to 24 categories extracted from the guidance documents, including 11 related to study design and description and 13 related to study outcomes. Overall, a weak positive trend was observed over time (P=0.004, R^2=0.07, slope=0.63%/year), with overall adherence (the sum of study descriptors and outcomes) ranging

---

*Corresponding author. lemauck@mail.med.upenn.edu.

from 32±16% to 58±14% in any individual year. There was no impact of the publication of the guidance documents on adherence (P=0.264 to 0.50). Given that improved adherence would expedite translation, we discuss the reasons for poor adherence and outline approaches to increase and promote their more widespread adoption.

INTRODUCTION

Articular cartilage defects are common (1–4) and, left untreated, can lead to disabling joint disease (5). As such, there has been considerable focus on developing new treatments. The field of regenerative cartilage therapeutics has evolved substantially since the advent of bone marrow stimulation and abrasion arthroplasty techniques (6, 7), to include both autograft and allograft osteochondral transplantation (8, 9) and the widespread use of cell-based autologous chondrocyte transplantation (ACI) (10). Likewise, regenerative medicine–based approaches have been developed to improve on these techniques [for review, see (11)], including in vitro–grown engineered constructs (12–14), biomaterial-based cell delivery systems (for next generation ACI), and materials that serve as adjuvants to repair (used in conjunction with microfracture), with several "first in human" clinical trials of new repair technologies reported over the last few years (15–17). The appearance of these reports speaks to the intense innovation occurring within this domain to provide both functional and durable repair of cartilage injuries.

In most (if not all) cases, translation of emerging regenerative approaches to clinical practice involves preclinical evaluation in an animal model. Large animal models (in particular equine, ovine, caprine, porcine, and canine) are commonly used for final (pivotal) preclinical studies. Critical evaluation of new technologies in large animals can highlight the safety and efficacy of new therapies, and also redirect designs to improve efficacy when preliminary versions fail. A number of recent reviews outline the strengths and weaknesses of these various large animal repair models (18–21).

Beyond selecting a species in which to perform the pivotal trial, investigators must carry out a well-performed study that will be accepted by their peer community and the regulatory agencies that will consider data generated toward the initiation of a human clinical trial. To that end, several resources exist to orient investigators to best practices, including guidance documents published by governing agencies including the U.S. Food and Drug Administration (FDA, first draft 2007) (22) and the European Medicines Agency (EMA, first draft 2008) (23) as well as expert-panel reports by the American Society for Testing and Materials (ASTM International, first draft 2005) (24) and the International Cartilage Repair Society (ICRS) (20). These guidance documents provide specific details related to study design, desired time points, and endpoint analyses to guide the proper execution of the study (Table S1 and Table S2). In each document, however, there is room for interpretation, acknowledging ongoing developments in the field and the difficult decisions that are made in carrying out such studies. For example, the language on an appropriate defect location is vague, recommending that investigators choose an "orthotopic" (23) or "similar to human" (22) site (see Table S1). Furthermore, these documents provide guidance but are not *per se*

mandates, and so investigators can make their own decisions as to the appropriate study design and the type and number of outcome assays employed.

Given the maturation of the field of cartilage repair as well as the recent publication of these guidance documents, we were curious as to the extent to which the field had incorporated these recommendations into their work. To answer this question, we performed a systematic review inclusive of the most common large animal models, capturing data from full-length peer-reviewed publications over the last two decades (1994–2014, see Supplementary publication listing). We then reviewed the extant literature that met our inclusion criteria (studies reporting on repair of chondral defects in large animal species) and used quantitative analysis to determine the degree to which these studies adhered to the guidance document recommendations.

Our overall hypothesis was that the field as a whole would show a quantifiable improvement in adherence to the guidance documents over the time period considered. Given that these documents appeared during the period of our review (for example, 2005, 2007, and 2008 for ASTM, FDA, and EMA, respectively), we postulated that there would be an inflection in the trajectory of adherence, with greater adherence after the publication of the guidance documents. We further expected that those studies with longer-term time points and those using larger species (both of which would increase study costs) would show a greater adherence to these published criteria, as these would more likely represent pivotal rather than preliminary studies.

RESULTS

The initial review of the extant guidance documents defined categories for subsequent analysis. These are defined in Table S1 for study descriptors and Table S2 for study outcomes, according to the governing agency defining the standard. Those agencies that focus on standardization of measures (for example, ASTM) provide the greatest detail in terms of execution of outcome assays, whereas other publications (for example, EMA) were intended as guidelines and so provide less-specific details (Fig. 1).

Next, we used our established inclusion/exclusion criteria to identify peer-reviewed studies for consideration. Our search identified 1187 articles on cartilage repair in large animals published as full-length manuscripts over the last 20 years (see Materials and Methods). Application of exclusion criteria reduced the total number to 123 for analysis. Of these, 9 were not eligible upon detailed review. Thus, 114 studies were included in this analysis, of which 20 were conducted in horses, 23 in sheep, 18 in goats, 41 in (mini) pigs, and 12 in dogs (Fig. 2).

All studies were reviewed and scored by one of the three reviewers. When questions arose, each of the three graders reviewed and discussed the publication, and a consensus score was assigned. Scores for all studies considered (columns) and their degree of adherence to each of the 24 categories (rows) was visualized by color-mapping the resulting scores for each study (Fig. 3). These were further analyzed by calculating the percentage of studies reporting on each study descriptor and study outcome.

The number of published large animal cartilage repair studies ranged from 1 per year to a peak of 16 in 2009 (Fig. 4A). Overall adherence to guidance documents (that is, overall assessment) showed a weak positive trend with time (P=0.004, R^2=0.07, slope=0.63%/year, Fig. 4B), ranging from 32±16% to 58±14% for a single year. However, there was no effect of the publication of the guidance documents on adherence (documents published in 2005, 2007, and 2008). That is, the effect of time on the level of adherence did not change in the period following any one of these publications (P=0.26 to 0.50). In fact, the slope of adherence versus time decreased after issuance of the guidance documents (0.89 to 1.11%/year before publication versus 0.07 to 0.48 after publication). Since 2003, most published studies had an overall adherence of ~50%.

We also considered adherence in each subset of categories, study descriptors (11 categories) and study outcomes (13 categories). Not surprisingly, this analysis showed that adherence to study descriptors was, without exception, higher than adherence to study outcomes (Fig. 4C). The only category not well covered in the subset of study descriptors was relative lesion size (<5% of all studies). This is an important category on which to report given that relatively small lesions may occupy a significant portion of the joint in some species (25). Of the 13 study outcomes, histological analysis was most commonly reported, with 97% of studies reporting on this category. It should be noted, however, that there was little consensus on the histological methodology employed, with some studies simply reporting very basic histology and others performing multiple stains and visualization by advanced imaging modalities (for example, polarized light). Likewise, some studies provided semi-quantitative scoring of histological outcomes, whereas others did not, and those that did provide semi-quantitation used a number of different scoring systems. Of the remaining study outcomes, only gross view (80%) and defect fill (60%) were reported by >50% of the studies examined. As with histology, a variety of qualitative and semi-quantitative assessments were employed. All other study outcomes fell below 50% adherence, with follow-up arthroscopy (reported by <8%) and gene expression (reported by 11%) being the two least reported.

When these metrics were analyzed across species, a similar pattern was observed (Fig. 4D). In general, findings showed that for overall adherence, equine > canine > ovine = caprine > porcine. For all species, study descriptors showed a higher level of adherence than the study outcomes. Somewhat in keeping with one of our initial hypotheses, studies that used horses showed a slightly higher score in terms of overall adherence (59±12%) and study outcomes (50±20%) compared to other species. These differences were only significant in comparison to (mini-)pigs, however, where overall adherence was 49±10% (P<0.001 versus equine) and study outcome adherence was 30±12% (P<0.01 versus equine). No other differences were observed among species.

Another of our original hypotheses was that studies that were longer or larger would have better adherence. Contrary to this hypothesis, the correlation between duration and overall adherence was 0.27 (with a P value of 0.003), indicating a weak correlation (Fig. 4E). Likewise, the correlation relating the number of animals used in a study to overall adherence was 0.001 (P = 0.994), suggesting no correlation existed between these factors (Fig. 4F). Given the labor and cost involved in studies using a large number of animals for a long

duration, it was surprising to find such weak correlations. Additional data comparing numbers of animals used, unilateral versus bilateral surgeries, numbers of groups per study, samples per group, and defects per joint across species are presented in the Supplementary Materials.

DISCUSSION

Activity in the field of cartilage repair and regeneration is well evidenced by the increase in number of translational cartilage repair studies over the last two decades (Fig. 4A) as well as several recent "first-in-human" reports. Guidance documents published by U.S. and European regulatory agencies provide direction for performing animal studies (20, 22–24), a framework in which to validate safety and efficacy prior to human trials, and should enable better comparison between studies to expedite progress in the field. Given that the first of these guidance documents was published ~2005, our overriding hypothesis was that the field as a whole would show a quantifiable improvement in reporting on the study design and outcome criteria outlined in these documents. To assess this, we performed a formal systematic review and also extracted quantifiable adherence data from the past two decades of large animal research in terms of study descriptors and study outcomes. In doing so, we tested the hypothesis that there would be an inflection in the trajectory of adherence to these guidance documents based on the timing of their publication. On the basis of our experience with such model systems, we further expected to find that those studies with longer-term time points and those that used larger species and a greater number of animals (both of which would increase study costs) would show a greater adherence to these published criteria.

Unfortunately, our initial hypotheses were generally not supported by the data extracted from the literature. Although there has been a slow increase in overall adherence to the guidance criteria over the past two decades (with recent studies reaching 50 to 60%), there was little impact (that is, change in trajectory) as a consequence of the publication of the guidance documents. Even more surprising, there seemed to be only weak or nonexistent correlations between overall adherence and the length of the study or cost of the study animals used. On the basis of the publications analyzed and the methodology used in this study, it is clear that the field has not responded to the publication of the guidance documents, and so there is little homogeneity in the reporting of these studies and the outcome assays that are being employed.

To probe the data set a bit further, we subdivided the analysis into study descriptors and study outcomes, in which the former enumerated basic information regarding the animals and surgical procedures and the latter summed the number of outcome assays used with respect to those suggested by the guidance documents. This analysis showed that the study descriptor categories were in general met at a rate of ~75%. It was surprising that the rate of adherence was not higher, and suggested to us that there is simply a lack of standardization in the field in terms of reporting on these parameters. Improvement in adherence in these categories could be achieved if authors simply reported on the age, weight, and gender of the animals used; these last two parameters were reported by <50% of the studies analyzed but can have a major influence on cartilage repair potential (in humans) (26–28). In addition, in

a complicated study design, the overall study is best understood when visualized in a flow chart, as is commonly used in clinical studies (29). The above are standard components of any animal protocol, but are not generally included in the extant literature. One recommendation for the field might be to better adhere to the ARRIVE (Animals in research: Reporting in vivo experiments) guidelines for publication (30), with select components of approved animal protocols (such as flow diagrams, tables of weights, gender, groups for analysis) provided with each published animal study (as supplementary material), so that the full complement of information is easily accessible to the community while not making published methods burdensomely long.

Some descriptors that directly influence experimental outcomes were reported on but done so with insufficient detail so as to enable replication. For instance, most studies reported some aspect of post-surgical rehabilitation, but few provided detailed descriptions (with the exception of those studies performed in horses), despite this having a profound impact on outcomes in both humans and animals (31). In addition, the description of the lesion depth and type was often vague or missing and did not always specifically differentiate between partial chondral, full chondral, and osteochondral defects [as described by Cucchiarini *et al.* (32)]. If the technique and visualization allows, the description of the defect creation should also carefully describe whether the calcified cartilage was removed or not, and if bleeding into the defect was present–again, these are rarely noted in published studies despite their impact on regeneration.

When it came to study outcomes, adherence to the guidance documents was by far less robust, with most studies falling in the range of ~40% adherence. Indeed, only three categories (histology, gross visualization, and assessment of defect fill) were reported on by more than 50% of the studies we analyzed. Many other important categories (for example, biomechanics and biochemistry of the regenerate tissue) were reported on by fewer than 25% of these studies. In some categories, such as relative defect size and gene expression, only one of the documents recommended reporting on this parameter, and so low adherence here might reflect an overall lack of consensus as to the value of such outcomes. Alternatively, some assays, such as follow up arthroscopy, are not possible in all species, and so poor adherence in this category might reflect the technical challenges inherent to some species. Even with these caveats, however, our findings suggest that there is very little consensus in the field as to the most important study outcomes to measure and report in these large animal models.

DO AS THEY SAY, NOT AS WE DO

Although it seems logical to follow the guidelines set forth by regulatory agencies ("do as they say"), researchers appear to largely ignore these recommendations ("not as we do"). In fact, the gulf between what they say we should do and what we actually do is quite large. Subjecting our own recent large animal model of cartilage repair (33) to this same analysis shows that our study fares no better than the industry standard. For our recent cartilage repair study in minipigs, we achieved an overall adherence of only 56%; whereas our adherence to study descriptors was a respectable 73%, our adherence to study outcomes was a lowly 42%. Given that adherence to study descriptors are generally at an acceptable level

and because improvements in this area could be easily made, we will focus our recommendations on study outcomes, in which a change in approach will be required for significant improvements to be made across the field.

There are several potential reasons for a lack of adherence in terms of study outcomes. First and foremost, not every study group has access to the infrastructure and expertise needed to conduct certain measurements. Indeed, those outcomes that are relatively easy and inexpensive to perform (histology and gross view imaging) are well covered across studies, whereas those requiring specialized expertise (for example, mechanical testing) and equipment (for example, follow-up MRI) are underrepresented. A solution to this infrastructure and expertise problem might be build better collaborations and research networks, or to create designated federally funded core facilities for the performance of these more advanced assays.

A second (and simple) reason for the low adherence rate might be that not every study has the same intent (for example, some being considered to be "pilot" and others "pivotal"). In pilot studies [including our own (33)], one naturally focuses on only one or a few outcomes within the expertise of the particular group and most predictive (in the opinion of the group) for what the experimental groups intend to pursue for further study. Such pilot studies are critical steps in advancing any technology and certainly need to be performed. However, it is still crucial to extract as much useful information as possible by increasing the number of outcome measurements obtained from each specimen in a non-destructive manner. Further, development and use of tools for repeated measures (for example, MRI, arthroscopy) on the same animals might improve the depth of reported outcomes while not overly increasing costs.

Another consideration is that different cohorts (basic scientists, engineers, clinicians) likely have differing opinions as to which of the outcomes is most important. The guidance documents are under constant review and improvement, and new insights from the field are continually incorporated (34). Given that so many outcome measures are simply not reported at present, it might well be that these documents should be refined further to develop a minimal set of recommended outcome measurements that is agreed upon (and followed) by all stakeholders.

A final, and perhaps the most important, factor defining the overall level of adherence is of course the amount and source of funding. Large animal trials are expensive, and not every group has the resources to analyze the full set of recommended outcomes, even though the marginal costs for performing additional assays might be small compared to the animal study cost. In some cases, funding might be available only for a pilot trial, whereas in other cases, funding might be available for a pivotal study with the intent of translating the findings to humans. This is not to say that pilot trials shouldn't be done. Rather, these studies should be performed and defined, perhaps with a different set of criteria for moving technology to the next and more rigorous (pivotal) level. An additional complication is that when studies are industry-sponsored, there is the added element of financial pressure, and not every industry-sponsored study is intended for open publication in the literature (limiting our ability to capture these data). Here, the need to commercialize a product as soon as

possible might set objectives at the level of providing just enough evidence to regulatory bodies (but not to the general public) so as to meet the bar for market admittance and initiation of a human clinical phase I trial. This is in no way judgment, but rather speaks to the reality of the world we live in, wherein timing, competitive advantage, and economy might mean the difference between market success and bankruptcy.

To address these deficiencies, our summary recommendation for those performing large animal studies is to "do as they say, not as we do". The guidance documents are quite robust, and all suggested outcomes have value and will accelerate progress if more routinely employed. The community should convene and come to consensus on definitions of minimal requirements that identify pilot and pivotal studies, and both editorial and regulatory bodies should enforce these criteria in advancing publications or new products. To make this possible and increase adherence overall, regulatory and funding agencies should provide additional resources to enable this more rigorous transition to translation. Coordinated effort and funding are necessary to improve outcomes and expedite the development of new technologies that can provide functional repair for the many patients suffering from articular cartilage injuries.

MATERIALS AND METHODS

Study design

For this systematic review, guidelines from three agencies were first synthesized to develop categories by which to assess the literature. Next, the literature on animal models to study cartilage repair were systematically reviewed and scored based on the adherence to these categories (search methodology and inclusion/exclusion criteria are described below). Data generated were then analyzed statistically to assess the impact of the publication of the guidance documents on whether these categories were reported on by the identified studies, as well as to assess species-specific differences in methods and reporting standards.

Synthesis of recommendations and categories for assessment

Three guidelines were considered: 1) the US Food and Drug Administration (FDA) Guidance for Industry: Preparation of IDEs and INDs for Products Intended to Repair or Replace Knee Cartilage (last accessed 4/30/2014) (22), 2) the European Medicines Agency (EMA) Guideline on Human Cell-based Medicinal Products (23) with International Standard ISO/EN 10993 (35) (last accessed 4/30/2014) and 3) the American Society for Testing and Materials (ASTM) International F2451-05 (2010) Standard Guide for in vivo Assessment of Implantable Devices Intended to Repair Articular Cartilage (24). Recommendations from these documents (Table S1 and Table S2) were sorted as 'study descriptors' (Table S1) and 'study outcomes' (Table S2). A total of a 24 categories were identified (11 study descriptors and 13 study outcomes).

Literature search and scoring

The literature search, screening, and eligibility testing followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (29). To retrieve published studies employing the five most common large animal models (horse, sheep, goat,

(mini-) pig, and dog) we searched the PubMed database (www.ncbi.nlm.nih.gov/pubmed) and the official journal of the International Cartilage Repair Society (ICRS) "Cartilage" on April 18th 2014, using the search term "cartilage_ specific species name_repair" (see Fig. 2). The two primary inclusion criteria included that the study be in a large animal (dog, pig, goat, sheep, horse), that the defect be chondral only (that is, not including the subchondral bone). The former represented studies that were testing a mature technology in valid pre-clinical model, and the latter was enforced to focus (and restrict) our analysis to cartilage repair apporaches (rather than appoaches seeking to repair both cartilage and bone at the same time). After screening articles for inclusion and exclusion criteria and eligibility, a total of 114 studies were identified and reviewed for adherence. If a category was fully reported, the study was credited one point. If the category was not reported/measured, a score of zero was assigned. If a category was partially reported, a score of 0.5 was assigned. Eligibility and scoring were reconciled by group discussion when not clear (i.e., all authors read the manuscript in question and weighed in to assign the score). Points per study were calculated as percentages of "overall adherence" (all 24 categories included), "study descriptors" (11 categories), or "study outcomes" (13 categories).

Statistical analysis

Statistical analysis was performed using SPSS® (Version 20, IBM, NY, USA). Pearson's correlation coefficients were computed for study duration versus overall adherence ($\alpha=0.05$). ANOVA with Bonferroni post-hoc testing ($\alpha=0.05$, two-tailed) was used to make comparisons for species specific statistics after establishing normality of data (Kolmogorov-Smirnov test). To assess whether individual guidance documents impacted adherence, ANCOVAS were performed, with a separate analysis for each document (i.e., using years 2005, 2007, or 2008 as the grouping variable) using time as a co-variate ($\alpha=0.05$, two-tailed). The interaction term between group and time was used to indicate if a particular document significantly impacted adherence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Curl WW, Krome J, Gordon ES, Rushing J, Smith BP, Poehling GG. Cartilage injuries: a review of 31,516 knee arthroscopies. Arthroscopy. 1997; 13:456–460. [PubMed: 9276052]

2. Widuchowski W, Widuchowski J, Trzaska T. Articular cartilage defects: study of 25,124 knee arthroscopies. Knee. 2007; 14:177–182. [PubMed: 17428666]

3. Hjelle K, Solheim E, Strand T, Muri R, Brittberg M. Articular cartilage defects in 1,000 knee arthroscopies. Arthroscopy. 2002; 18:730–734. [PubMed: 12209430]

4. Hintermann B, Boss A, Schafer D. Arthroscopic findings in patients with chronic ankle instability. Am. J. Sports Med. 2002; 30:402–409. [PubMed: 12016082]

5. Heir S, Nerhus TK, Rotterud JH, Loken S, Ekeland A, Engebretsen L, Aroen A. Focal cartilage defects in the knee impair quality of life as much as severe osteoarthritis: a comparison of knee injury and osteoarthritis outcome score in 4 patient categories scheduled for knee surgery. Am. J. Sports Med. 2010; 38:231–237. [PubMed: 20042546]

6. Pridie K. A method of resurfacing osteoarthritic knee joints. J. Bone Joint Surg. [Br]. 1959; 41:618–619.

7. Steadman JR, Rodkey WG, Rodrigo JJ. Microfracture: surgical technique and rehabilitation to treat chondral defects. Clin. Orthop. Relat. Res. 2001:S362–369. [PubMed: 11603719]

8. Alford JW, Cole BJ. Cartilage restoration, part 2techniques, outcomes and future directions. Am. J. Sports Med. 2005; 33:443–460. [PubMed: 15716263]

9. Chahal J, Gross AE, Gross C, Mall N, Dwyer T, Chahal A, Whelan DB, Cole BJ. Outcomes of osteochondral allograft transplantation in the knee. Arthroscopy. 2013; 29:575–588. [PubMed: 23544690]

10. Brittberg M, Lindahl A, Nilsson A, Ohlsson C, Isaksson O, Peterson L. Treatment of deep cartilage defects in the knee with autologous chondrocyte transplantation. N. Engl. J. Med. 1994; 331:889–895. [PubMed: 8078550]

11. Johnstone B, Alini M, Cucchiarini M, Dodge GR, Eglin D, Guilak F, Madry H, Mata A, Mauck RL, Semino CE, Stoddart MJ. Tissue engineering for articular cartilage repair – the state of the art. Eur. Cell. Mater. 2013; 25:248–267. [PubMed: 23636950]

12. Erickson IE, Kestle SR, Zellars KH, Farrell MJ, Kim M, Burdick JA, Mauck RL. High mesenchymal stem cell seeding densities in hyaluronic acid hydrogels produce engineered cartilage with native tissue properties. Acta. Biomater. 2012; 8:3027–3034. [PubMed: 22546516]

13. Peng G, McNary SM, Athanasiou KA, Reddi AH. Surface Zone Articular Chondrocytes Modulate the Bulk and Surface Mechanical Properties of the Tissue-Engineered Cartilage. Tissue Eng. Part A. 2014

14. Kelly TA, Roach BL, Weidner ZD, Mackenzie-Smith CR, O'Connell GD, Lima EG, Stoker AM, Cook JL, Ateshian GA, Hung CT. Tissue-engineered articular cartilage exhibits tension-compression nonlinearity reminiscent of the native cartilage. J. Biomech. 2013; 46:1784–1791. [PubMed: 23791084]

15. Sharma B, Fermanian S, Gibson M, Unterman S, Herzka DA, Cascio B, Coburn J, Hui AY, Marcus N, Gold GE, Elisseeff JH. Human cartilage repair with a photoreactive adhesive-hydrogel composite. Sci. Transl. Med. 2013; 5:167ra166.

16. Adachi N, Ochi M, Deie M, Nakamae A, Kamei G, Uchio Y, Iwasa J. Implantation of tissue-engineered cartilage-like tissue for the treatment for full-thickness cartilage defects of the knee. Knee Surg. Sports Traumatol. Arthrosc. 2014; 22:1241–1248. [PubMed: 23666378]

17. Saris D, Price A, Widuchowski W, Bertrand-Marchand M, Caron J, Drogset JO, Emans P, Podskubka A, Tsuchida A, Kili S, Levine D, Brittberg M. Matrix-Applied Characterized Autologous Cultured Chondrocytes Versus Microfracture: Two-Year Follow-up of a Prospective Randomized Trial. Am. J. Sports Med. 2014; 42:1384–1394. [PubMed: 24714783]

18. Ahern BJ, Parvizi J, Boston R, Schaer TP. Preclinical animal models in single site cartilage defect testing: a systematic review. Osteoarthritis Cartilage. 2009; 17:705–713. [PubMed: 19101179]

19. Chu CR, Szczodry M, Bruno S. Animal models for cartilage regeneration and repair. Tissue Eng. Part B Rev. 2010; 16:105–115. [PubMed: 19831641]

20. Hurtig MB, Buschmann MD, Fortier LA, Hoemann CD, Hunziker EB, Jurvelin JS, Mainil-Varlet P, McIlwraith CW, Sah RL, Whiteside RA. Preclinical Studies for Cartilage Repair: Recommendations from the International Cartilage Repair Society. Cartilage. 2011; 2:137–152. [PubMed: 26069576]

21. Cook JL, Hung CT, Kuroki K, Stoker AM, Cook CR, Pfeiffer FM, Sherman SL, Stannard JP. Animal models of cartilage repair. Bone Joint Res. 2014; 3:89–94. [PubMed: 24695750]

22. No authors listed, US Food and Drug Administration (FDA). Guidance for industry: Preparation of IDEs and INDs for products intended to repair or replace knee cartilage. 2011. http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

23. European Medicines Agency (EMA). Guideline on Human Cell-based Medicinal Products. 2008. www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003898.pdf

24. American Society for Testing and Materials (ASTM). ASTM F2451 - 05 Standard Guide for in vivo Assessment of Implantable Devices Intended to Repair or Regenerate Articular Cartilage. 2010. www.astm.org/Standards/F2451.htm

25. Wall EJ, Vourazeris J, Myer GD, Emery KH, Divine JG, Nick TG, Hewett TE. The healing potential of stable juvenile osteochondritis dissecans knee lesions. J. Bone Joint Surg. Am. 2008; 90:2655–2664. [PubMed: 19047711]

26. Smeriglio P, Lai JH, Dhulipala L, Behn AW, Goodman SB, Smith RL, Maloney W, Yang F, Bhutani N. Comparative Potential of Juvenile and Adult Human Articular Chondrocytes for Cartilage Tissue Formation in 3D Biomimetic Hydrogels. Tissue Eng. Part A. 2014

27. Lee HY, Han L, Roughley PJ, Grodzinsky AJ, Ortiz C. Age-related nanostructural and nanomechanical changes of individual human cartilage aggrecan monomers and their glycosaminoglycan side chains. J. Struct. Biol. 2013; 181:264–273. [PubMed: 23270863]

28. Kreuz PC, Muller S, von Keudell A, Tischer T, Kaps C, Niemeyer P, Erggelet C. Influence of sex on the outcome of autologous chondrocyte implantation in chondral defects of the knee. Am. J. Sports Med. 2013; 41:1541–1548. [PubMed: 23729686]

29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6:e1000097. [PubMed: 19621072]

30. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol. 2010; 8:e1000412. [PubMed: 20613859]

31. Vogt S, Angele P, Arnold M, Brehme K, Cotic M, Haasper C, Hinterwimmer S, Imhoff AB, Petersen W, Salzmann G, Steinwachs M, Venjakob A, Mayr HO. Practice in rehabilitation after cartilage therapy: an expert survey. Arch. Orthop. Trauma Surg. 2013; 133:311–320. [PubMed: 23263155]

32. Cucchiarini M, Madry H, Guilak F, Stoddart MJ, Koon Wong M, Roughley P. A vision on the future of articular cartilage repair. Eur. Cell. Mater. 2014; 27:12–16. [PubMed: 24802612]

33. Fisher MB, Belkin NS, Milby AH, Henning EA, Bostrom M, Kim M, Pfeifer C, Meloni G, Dodge GR, Burdick JA, Schaer TP, Steinberg DR, Mauck RL. Cartilage Repair and Subchondral Bone Remodeling in Response to Focal Lesions in a Mini-Pig Model: Implications for Tissue Engineering. Tissue Eng. Part A. 2014

34. Sah RL, Ratcliffe A. Translational models for musculoskeletal tissue engineering and regenerative medicine. Tissue Eng. Part B Rev. 2010; 16:1–3. [PubMed: 19905871]

35. International Organisation for Standardization (ISO). ISO/TC 194: Biological and clinical evaluation of Medical Devices. 2009. www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_browse.htm

Fig. 1.
Recommendations from regulatory agencies for appropriate descriptors and outcomes for translational large animal studies in cartilage repair. Color scheme indicates whether the documents provide detailed (=green), loose (=yellow), or no (=red) recommendations in a category.

Database PubMed Central and official ICRS journal "Cartilage" searched on 4/18/2014
Search terms:

cartilage horse repair	→ 150 identified publications
cartilage sheep repair	→ 226 identified publications
cartilage goat repair	→ 114 identified publications
cartilage (mini-)pig repair	→ 368 identified publications
cartilage dog repair	→ 329 identified publications

screening for inclusion and exclusion criteria

|	
cartilage horse repair	→ 127 excluded
cartilage sheep repair	→ 203 excluded
cartilage goat repair	→ 94 excluded
cartilage (mini-)pig repair	→ 326 excluded
cartilage dog repair	→ 314 excluded

test for eligibility

cartilage horse repair → 3 additionally excluded (1 osteochondral defect,1 additional data from previously reported study, 1 published before 1994)
cartilage sheep repair → no additional exclusions
cartilage goat repair short → 2 additionally excluded (1 due to double published results, 1 used one term animal and reported on translational results in human)
cartilage (mini-)pig repair → 1 additionally excluded (osteochondral defect)
cartilage dog repair → 3 additionally excluded (2 not cartilage repair models, 1 due to double published results)

inclusion of studies for analysis

cartilage horse repair	→ 20 studies
cartilage sheep repair	→ 23 studies
cartilage goat repair	→ 18 studies
cartilage (mini-)pig repair	→ 41 studies
cartilage dog repair	→ 12 studies

Fig. 2.
Literature search, screening, and eligibility testing using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (29).
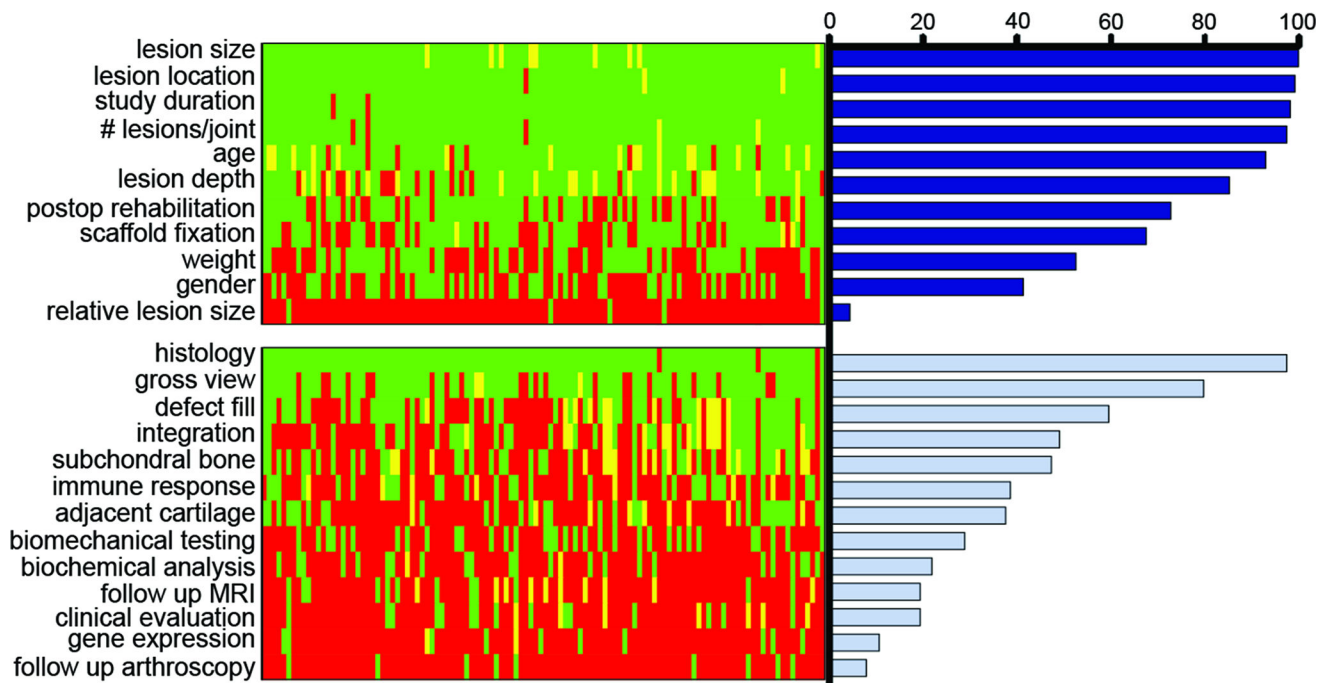
Fig. 3.
Array depiction and quantification of adherence to guidance document recommendations.
Each row indicates a different category (descriptors on top, outcomes on bottom), while each
column represents an individual study. Green indicates full adherence, red indicates non-
adherence, and yellow indicates partial adherence. Bars at right show average percentage
adherence across all studies in a given category.
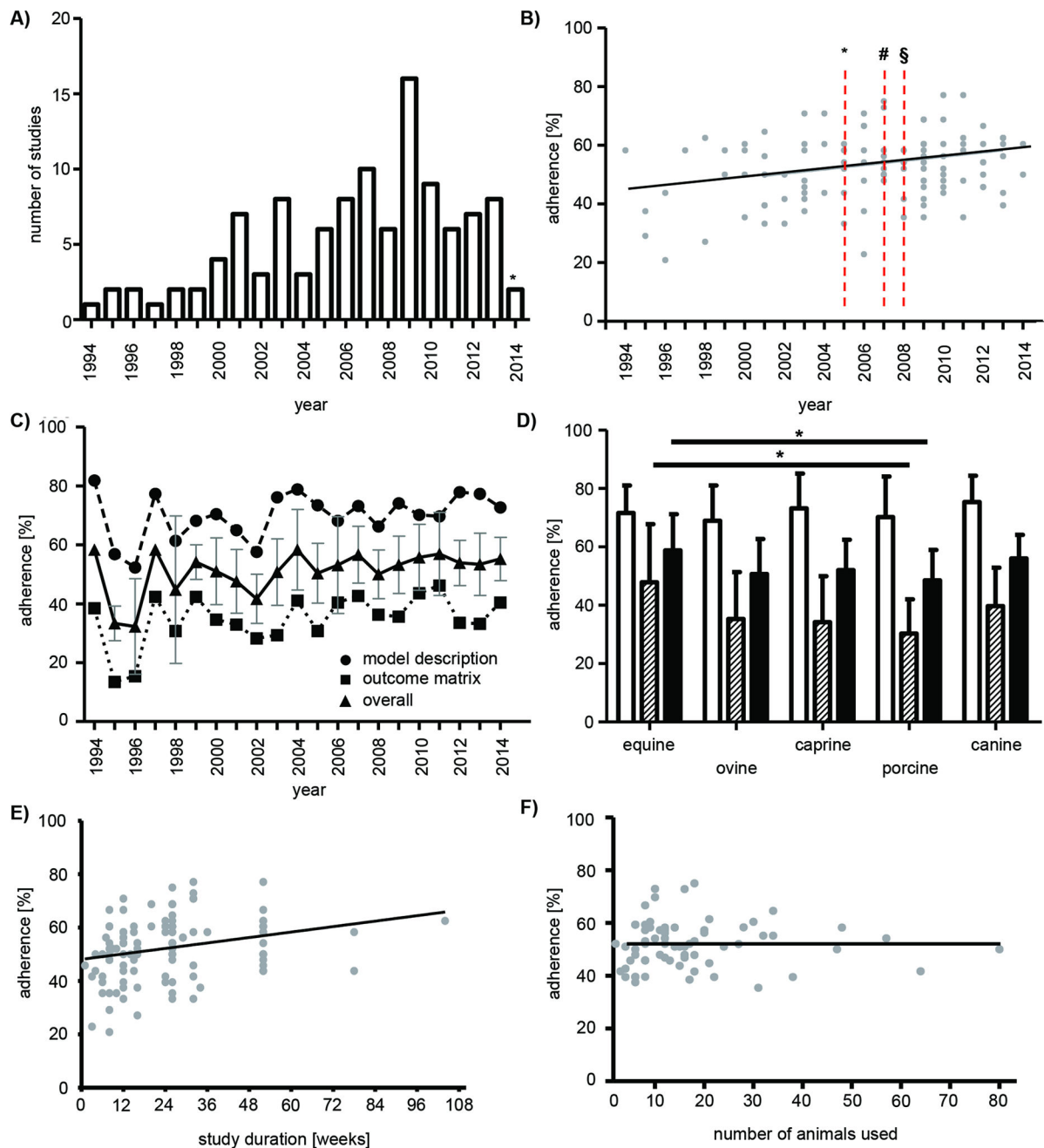
Fig. 4.

(**A**) Number of studies per year (*through 4/18/14). (**B**) Overall adherence by study and year. Vertical lines indicate the time of publication of the *ASTM, # FDA, and § EMA documents. Regression shows slight increase in overall adherence with time (p=0.004, R^2=0.07, slope=0.63%/year). **C**) Mean adherence per year, broken out into overall, study descriptors, and study outcomes. Study descriptors were reported at a higher level than study outcomes. Standard deviation is shown for overall adherence only. **D**) Adherence (in percent) across species (white bars = study descriptors, hatched bars = study outcomes, black bars = overall adherence, *P<0.05). **E**) Correlation between overall adherence and

study duration (R^2=0.075, P=0.003) and **F**) overall adherence and the number of animals (R^2<0.001, P=0.994).