Article

A Critical Look at Biomedical Journals' Policies on Animal Research by Use of a Novel Tool: The EXEMPLAR Scale

Ana Raquel Martins [1] and Nuno Henrique Franco [2,*]

[1] Faculty of Sciences, University of Porto, Rua do Campo Alegre S/N, 4169-007 Porto, Portugal;
    E-Mail: up201101588@fc.up.pt
[2] IBMC—Instituto de Biologia Molecular e Celular, University of Porto, Rua do Campo Alegre 823,
    4150-180 Porto, Portugal

* Author to whom correspondence should be addressed; E-Mail: nfranco@ibmc.up.pt;
    Tel.: +351-226-074-900 (ext. 2712).

Simple Summary: Biomedical journals have the responsibility to promote humane research. To gauge and evaluate journal policies on animal research, the *EXEMPLAR*—For "Excellence in Mandatory Policies on Animal Research"—scale is presented and applied to evaluate a sample of 170 biomedical journals, providing an overview of the current landscape of editorial policies on the ethical treatment of animals.

Abstract: Animal research is not only regulated by legislation but also by self-regulatory mechanisms within the scientific community, which include biomedical journals' policies on animal use. For editorial policies to meaningfully impact attitudes and practice, they must not only be put into effect by editors and reviewers, but also be set to high standards. We present a novel tool to classify journals' policies on animal use—the *EXEMPLAR* scale—as well as an analysis by this scale of 170 journals publishing studies on animal models of three human diseases: Amyotrophic Lateral Sclerosis, Type-1 Diabetes and Tuberculosis. Results show a much greater focus of editorial policies on regulatory compliance than on other domains, suggesting a transfer of journals' responsibilities to scientists, institutions and regulators. Scores were not found to vary with journals' impact factor, country of origin or antiquity, but were, however, significantly higher for open access journals, which may be a result of their greater exposure and consequent higher public scrutiny.

1. Introduction

Animal experimentation has, for centuries, not only been a cornerstone of scientifically-based biomedical research, but also the subject of heated academic and public debate [1]. Given the number and complexity of the arguments at play in this discussion, views on the moral acceptability of animal experiments are diverse, most of which fall in between the diametrically opposed views of those uncompromisingly against this practice and the more determined advocates of animal experimentation. Such diversity results from the interplay between such factors—among others—as the suffering and discomfort endured by animals [2–4], the purpose and validity of research [4–8], which species are used [6,9–11] or the degree of confidence in scientists and regulators [3,12,13], as well as the different positions on the weighting that should be given to each factor.

In spite of this multitude of positions, in western societies, the majority of the public tends to follow a moderate, mainstream view on the subject, according to which animal experimentation is seen as a legitimate and ethically acceptable scientific activity, provided it is guided by a scientifically sound and medically relevant objective, and that animal welfare is taken into consideration [8,13]. This "conditional approval" view is in line with the tenet of the Three Rs [14]—for the Replacement, Reduction and Refinement of animal use in science—widely accepted as a scientifically sound approach to help address the ethical dilemma put forth by animal research [8,15], as well as an overarching principle in current European legislation regulating animal use for scientific purposes [16].

The mandate of scientists to work with animals requires a continually renewed relationship of trust with society, whose views and expectations, on the other hand, legislators also take into account when establishing the legal framework within which scientists must work. This tacit "social license" granted to scientists warrants public trust in the relevance, competency, reliability, ethicality and transparency of animal experiments, which goes beyond the mere assumption of compliance with laws and regulations [17,18]. From this interplay results a tightly regulated environment for animal research, by both 'external' regulatory frameworks (such as legislation [16,19]) and self-regulatory initiatives from within the scientific community itself. These include peer review [20], training [21,22], the issuing of publication guidelines (such as ARRIVE [23] or the *Gold Standard Publication Checklist* [24]) or other scientific community-driven initiatives, such as the Basel Declaration for transparency in animal research [25]. Biomedical journals' policies also have great potential for improving principles and practice in animal research, since the motivation—and often the pressure—to publish may serve as an incentive for scientists to comply with journals' demands [26,27] on such issues as quality of research and reporting, or the ethical treatment of animals, particularly for high-impact journals.

Given the potential of scientific journals for promoting best practice in animal research, evaluating their level of concern for animal welfare and the quality of research becomes of the essence. In 2009, RSPCA's Nicola Osborne *et al.* [28] proposed a 12-item (each awarding one point) classifying scheme for scoring biomedical journals' policies on animal welfare and the Three Rs, and presented results of a

review of 236 biomedical journals' policies. They found that in 35% of their sample, animal use was not even contemplated in the guidelines for authors or anywhere else, in 18% animals were mentioned in some way but no perceptible guidelines were provided and the remaining scored rather poorly, with 37% scoring three or fewer points (out of 12).

Table 1. The Excellence in Editorial Mandatory Policies for Animal Research (EXEMPLAR) scale.

EXEMPLAR Scale		Score
A—Reporting of regulatory compliance		
Authors must attest of prior ethical approval of animal studies, or an analogous project evaluation process involving harm-benefit appraisal, (e.g., by providing documental evidence or ethics review board process reference) and include a statement of compliance with relevant legislation and national, international or institutional guidelines on animal care and use.		**5**
Authors must declare:	Ethical approval of studies or analogous evaluation process by competent authority, institutional animal care and use committee, animal welfare body or equivalent.	2
	Compliance with relevant national, international or institutional guidelines on animal care and use.	1
	Compliance with relevant legislation on the use of animals.	1
B—Quality of research and reporting of results		
Journal refers authors to relevant guidelines on the quality of reporting of animal studies (e.g., ARRIVE guidelines, Gold Standard Publication Checklist, or other)		**5**
Authors must include information regarding:	Animals (e.g., sex, age, genotype, background, supplier, acclimatization period, *etc.*)	1
	Experimental conditions (e.g., housing, lighting, temperature, feeding regime, environmental enrichment, *etc.*)	1
	Experimental design and statistics	1
	Experimental procedures and outcomes	1
C—Animal Welfare and ethics		
Journal demands that the methods described are coherent with best principles and practice on the ethical treatment of animals in research, e.g., by demanding strict compliance with relevant guidelines on animal care and use, the 3Rs for *replacement*, *reduction* and *refinement* or the journal's own policies.		**5**
Authors must:	state the rationale for choosing the animal model(s) used	1
	report the impact of the experiment on animal health and wellbeing	1
	describe any measures to minimize harm (e.g., anesthetics, analgesics, humane endpoints) and/or improve the wellbeing of animals (e.g., husbandry adaptations for more vulnerable animals)	1
	justify the necessity of any unrelieved pain, suffering or distress inflicted	1
D—Criteria for the exclusion of papers		
Meeting journal standards on animal ethics care and welfare is indispensable for manuscript acceptance and/or publication. Studies raising serious concerns over animal welfare, or presenting significant discrepancies between the approved protocol and methods described may be reported to the institution or committee responsible for ethical approval.		**5**
Studies raising serious ethical concerns (e.g., serious neglect regarding animal welfare or unjustifiable suffering considering the value of the experiment) may be rejected by editors or reviewers.		3
Journal states specific procedure(s) that will not be accepted for publication (e.g., use of muscle relaxants or paralytic drugs alone for surgery, severe lesion/trauma without anesthesia, or death as an endpoint)		1

The present study presents the "EXEMPLAR"—an acronym for "Excellence in Editorial Mandatory Policies for Animal Research"—Scale (Table 1) as a novel approach and tool to classify and evaluate scientific journals' editorial policies on animal use. The EXEMPLAR Scale aims to (a) gauge the consideration (or lack thereof) given by journals to the most relevant aspects of animal use in the life sciences; (b) present a novel way for scoring editorial policies and (c) propose a set of ideal standards for three key aspects in which journal policies can have an instrumental role in promoting best practice: regulatory compliance, quality of reporting and animal welfare and ethics; along with a fourth category for the enforcement of said policies.

This paper also presents an overview of the current landscape of biomedical journals' policies, by reviewing and classifying by the EXEMPLAR a sample of 170 journals publishing animal studies within three fields of biomedical research: Amyotrophic Lateral Sclerosis (ALS), Type 1 Diabetes (T1D) and Tuberculosis (TB). These fields were selected for being the subject of ongoing systematic reviews in our lab—which will later allow for a comparison between the principles reflected in explicit policies and the actual practice patent in the articles published by the journals in our sample—which in turn were selected for representing areas in which animal experimentation is recurrent and widespread.

2. Experimental Section

2.1. The Exemplar Scale

The EXEMPLAR scale (Table 1) was developed to classify scientific journals' editorial policies on animal studies, scoring them from zero to 20 points. The scale is divided into four categories, each of them regarding a relevant aspect of journal policies on animal research. The maximum score for each category is five points, granted to journals that uphold to the scale's gold-standard for that category. For any journal not abiding to the category's "gold-standard" criterion, up to four points can be attributed if it nevertheless abides to other relevant criteria listed for that category, each deemed a "standard" criterion.

Categories are divided as such: Category A—Regulatory compliance; Category B—Quality of research and reporting of results, Category C—Animal welfare and ethics; Category D—Criteria for exclusion of papers. Both Categories B and C have four standard criteria, each one worth one point. Category A has three standard criteria, with the item "authors must declare ethical approval of studies or analogous project evaluation process (…)" awarding two points. This is based on the assumptions that (a) project approval involves some level of harm-benefit appraisal by a third party—such as the competent authority or an ethics committee—and (b) that the latter is of greater significance for the humane treatment of animals than plain self-reported compliance with laws and legislation. Category D only has two standard criteria, one of which awarding three points to journals which state that papers may be reject upon ethical concerns. This also follows the rationale that some criteria are of greater relevance than others, in this case, the mere listing of procedures the journal is not willing to publish.

For reporting purposes, the overall EXEMPLAR score is presented along with a breakdown of the partial scores for each category, allowing for anyone knowledgeable of the scale to identify in which dimensions a given journal's policy is stronger or, on the other hand, where further development is due. For example, the hypothetical *EXEMPLAR score = **12**: (A-5; B-5; C-2; D-0)*—or ***12**:(5,5,2,0)*—would suggest that the journal in question follows "gold standard" criteria for both Categories A and B, that it fulfils half the criteria for Category C and that no policies regarding Category D are described.

2.2. Journals Search

A list of papers reporting studies on murine models of Type-1 Diabetes (T1D), Tuberculosis (TB) and Amyotrophic Lateral Sclerosis (ALS) published between 2011 and 2013 was retrieved by a ISI Web of Science™ (Core Collection v5.13.1) advanced search. The search was carried out in April 2014 using the queries *TS= (("NOD mouse" OR "NOD mice" OR "non obese diabetic" OR "nonobese diabetic") AND diabet*)*; *TS = ((mice OR mouse) AND tuberculosis)*; and *TS = ((mice OR mouse) SAME (ALS OR "amyotrophic lateral sclerosis"))*. The results were refined to only include articles in English reporting original research, and afterwards downloaded and archived as an ENDNOTE® database. After being refined, search results were as such: T1D: 655 papers; TB: 1107 papers; ALS: 1114 papers. The criterion for selecting which journals to classify was to only include those that had published three or more papers in each field over the time period selected (Figure 1). This resulted in a sample of 44 journals for T1D (which published 475/655, or 73%, of T1D papers retrieved); 65 journals for TB (which published 999/1107, or 72% of all TB papers retrieved) and 84 journals for ALS (which published 830/1114, or 75%, of all ALS papers retrieved).
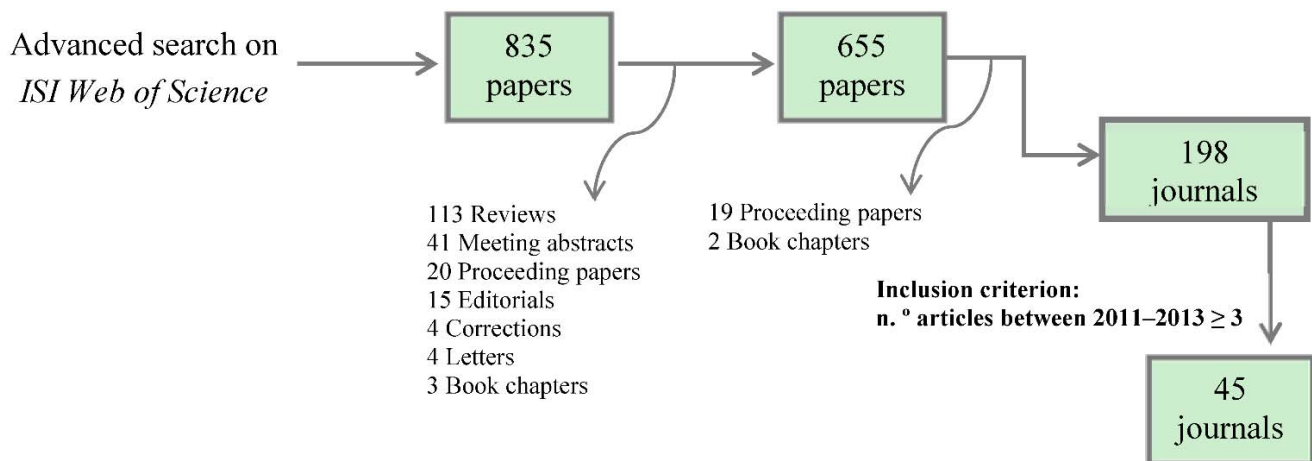

Figure 1. Schematic representation of the search and selection process, with the search for journals publishing studies on Type 1 Diabetes (T1D) being presented as an example. One of the journals from the T1D sample was later excluded because their editorial polices stated that it did not publish studies on animal models, leaving a final sample size of 44 journals.

2.3. Data Collection

Journal policies were collected by finding the "authors guidelines", "Instructions for authors" sections or equivalent on each journal's website, and saving them to a database. Whenever a journal required authors to comply with general policies of its publishing house, or subscribed to other guidelines (e.g., by the International Association of Veterinary Editors), these were also considered as a part of the journal's policies.

General information on each journal—impact factor, publisher, country and scientific category—were collected from the ISI Web of Knowledge Journal Citation Reports®, JCR Science Edition 2013, while information regarding the date (year) of the first issue and model of publication—*i.e.*, subscription based *vs.* open access—were retrieved from each journal's website. Only journals in which all published papers were readily and freely accessible without any embargo period were considered to be "open access".

All editorial policies were classified according to both the EXEMPLAR Scale (Table 1) and the classification scheme proposed by Osborne *et al.* [28].

2.4. Statistical Analysis

Chi-square tests were conducted to assess the relationship between the EXEMPLAR Score (ES) and qualitative variables (model of publication, country, and publisher), while for quantitative variables (impact factor, first year of edition and the score by the Osborne *et al.* scale) a logistic regression analysis was also performed. Linear relationships were assessed by the Pearson test. For some non-parametric statistical tests, the variable "ES" was transformed into a dichotomous variable, ES < eight and ES ≥ eight. A score of eight points was selected as a threshold for a minimally acceptable policy on animal use, since it can be awarded to a journal that complies with half the "standard" criteria defined for each category. Statistical analysis was conducted using SPSS Statistics 22 (IBM, Armonk, NY, USA; version 22.0).

3. Results and Discussion

3.1. Sample Characterization

The total sample comprised 170 journals, published by 54 academic publishers, with headquarters in 20 countries. Four publishers were responsible for publishing nearly half (49%) of the retrieved journals, namely Elsevier (36 journals), Wiley-Blackwell (22 journals), Springer (13 journals) and Nature Publishing Group (12 journals). Regarding country of origin, 95% of all journals were either based in the EU (81/170, 47 of which in the United Kingdom and 11 in the Netherlands) or the United States of America (USA) (80/170). The mean impact factor of the journals in this sample was 4.79 (Median = 3.63; Standard Deviation = 3.85). The median year of publication of first issue was 1987, with 75% of journals preceding 2002. Figure 2 represents the distribution of journals by field of research (of the three fields reviewed for this present study).
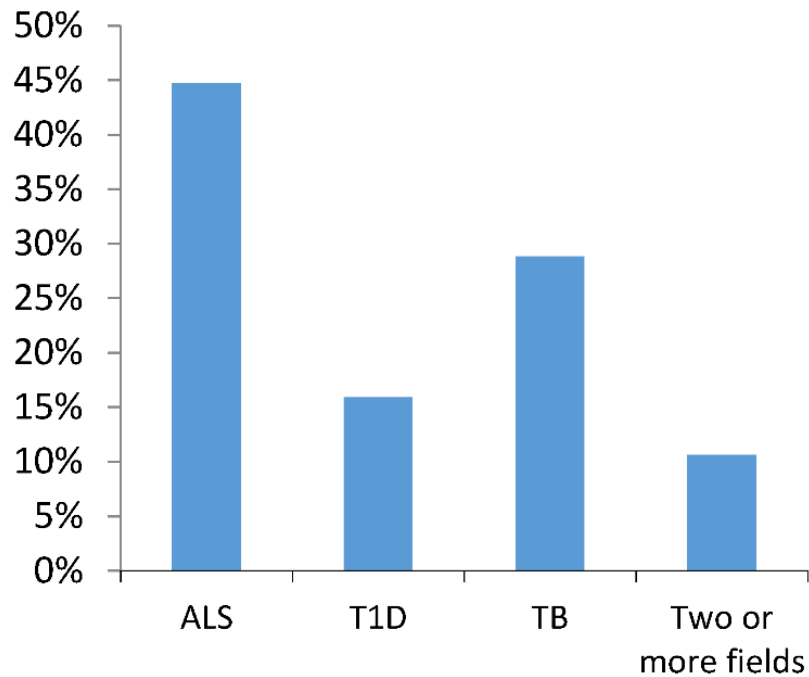
Figure 2. Topics covered by the journals in the sample (of the three selected). Journals publishing studies in more than one of the selected fields (n = 18) included two journals publishing papers on T1D and ALS, one journal publishing on Amyotrophic Lateral Sclerosis (ALS) and Tuberculosis (TB), ten journals publishing on both T1D and TB and five journals publishing papers on all three fields, between 2011–2013.

3.2. EXEMPLAR Score

No significant differences were found between the scores of the three sub-samples analyzed, the same happening for score distribution (Figure 3B). Results are hence presented for all journals pooled ($N = 170$, after removing duplicate entries from journals publishing in more than one field), unless stated otherwise. Category A ("reporting of regulatory compliance") registered the highest non-nil score of all categories (92%), with 5% of journals being awarded the top score of five points. As for Category B ("Quality of research and reporting of results"), 72% had a nil score. However, 18% of journals were awarded the top score for this category, virtually all of these for referring to the ARRIVE guidelines. As for Category C (Ethical treatment of animals) and Category D (Criteria for the exclusion of papers), 86% and 91% of journals scored zero points, respectively (Figure 4). Overall, only 18% (31/170) of journals scored eight points or higher, with the two top-scoring journals being published by the Public Library of Science (*PLOS*), namely *PLOS One*, with a maximum score of 20 points, and *PLOS Genetics*, which scored 15 points (the two other PLOS journals in the sample scored five and 10 points).
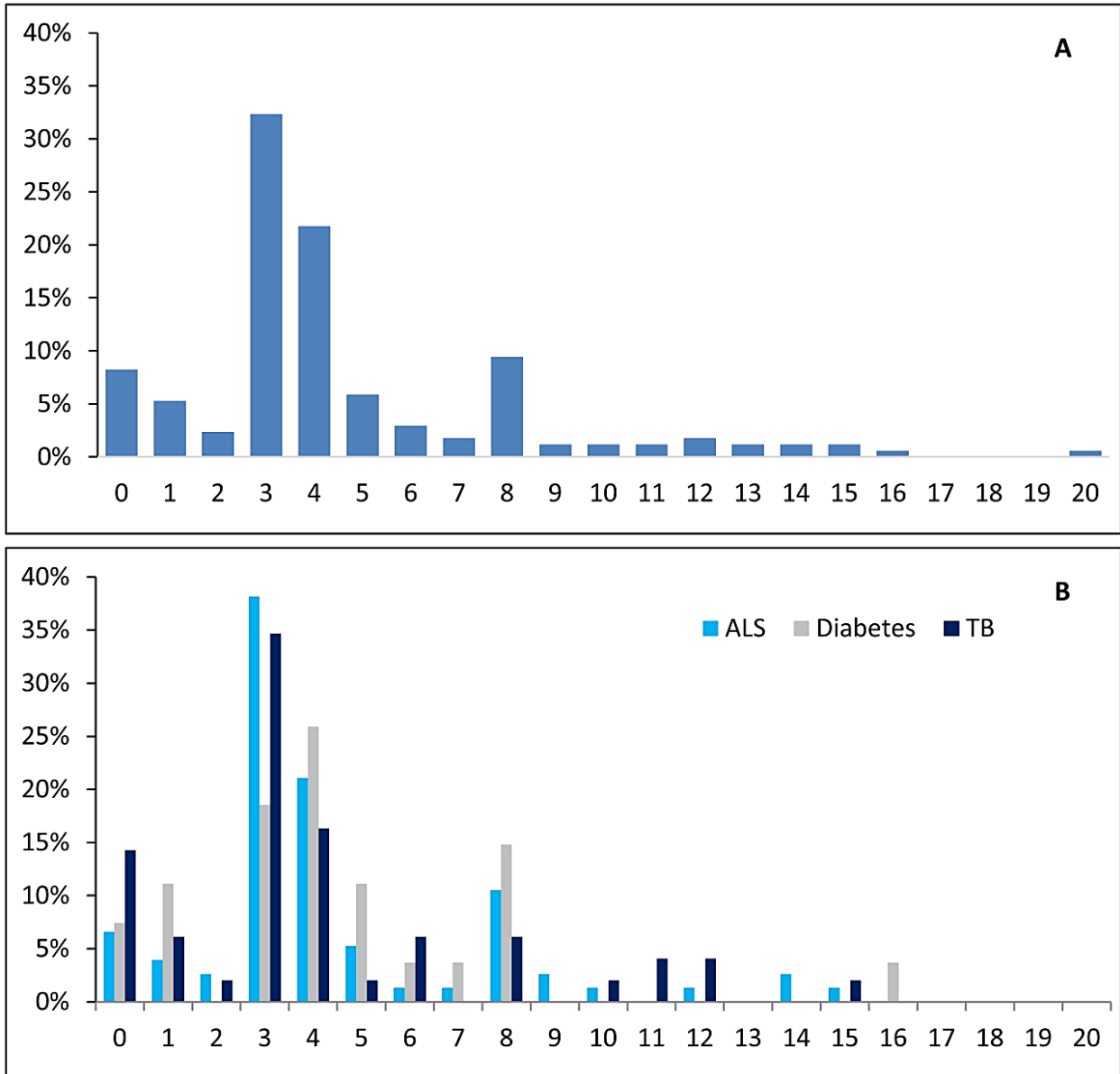
Figure 3. EXEMPLAR score distribution for the whole sample (Figure 3A; N = 170) and by field of research [Figure 3B, with 18 journals publishing studies in more than one field excluded; n(ALS) = 76; n(T1D) = 27; n(TB) = 49)]. The median score for the whole sample (N = 170) was 4 points, whereas the mean scores for ALS, T1D and TB were, respectively 4.45 (SD = 3.15), 4.52 (SD = 3.286) and 4.20 (SD = 3.54).
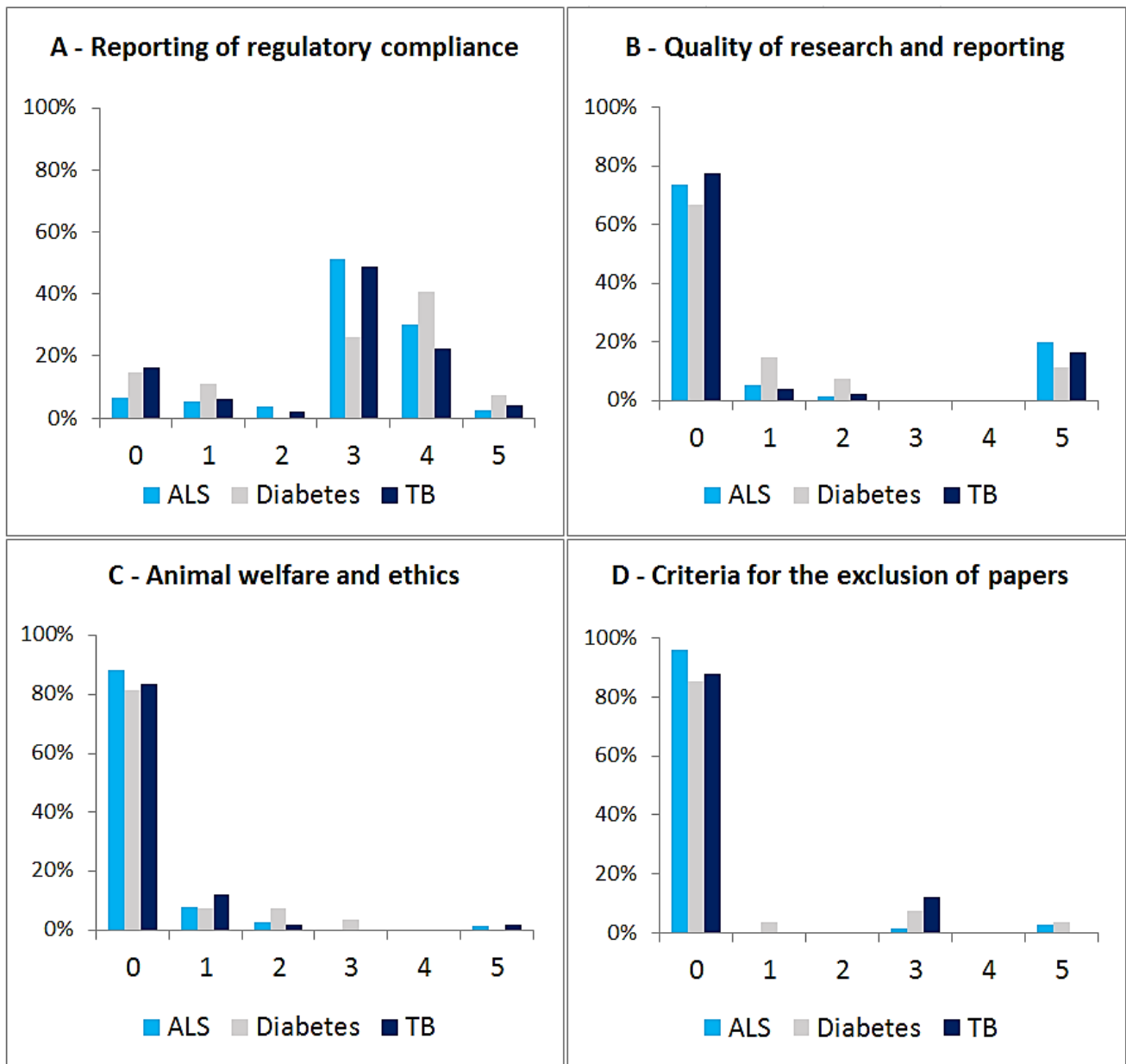
Figure 4. Scores for Categories **A**, **B**, **C** and **D**, for each. Journals publishing in more than one of the selected fields (n = 18) not shown [n(ALS) = 76; n(T1D) = 27; n(TB) = 49)].

3.3. Relationship between EXEMPLAR Score and Other Parameters

EXEMPLAR score and scores by the Osborne *et al.* scale (Figure 5) showed a positive correlation (Pearson's r = 0.623, *p* < 0.001). The sample showed a mean "Osborne score" of 2.98 points with a standard deviation of 1.28. Less than 25% of journals' scored more than three points (out of 12), and 8% of journals had a nil score by the Osborne *et al.* scale. Impact factor, country of origin, scientific category and first year of edition of scientific journals had no effect on the overall EXEMPLAR score. The model of publication (subscription *vs.* open access journals) was however shown to influence the overall EXEMPLAR score, as significantly more open access journals scored eight points or above ($\chi^2(1) = 17.87$, $p < 0.001$) than subscription based journals (Figure 6). Also, scores varied significantly between publishers ($\chi^2(53) = 96.968$, $p = 0.001$), as well as between journals from the same publisher.
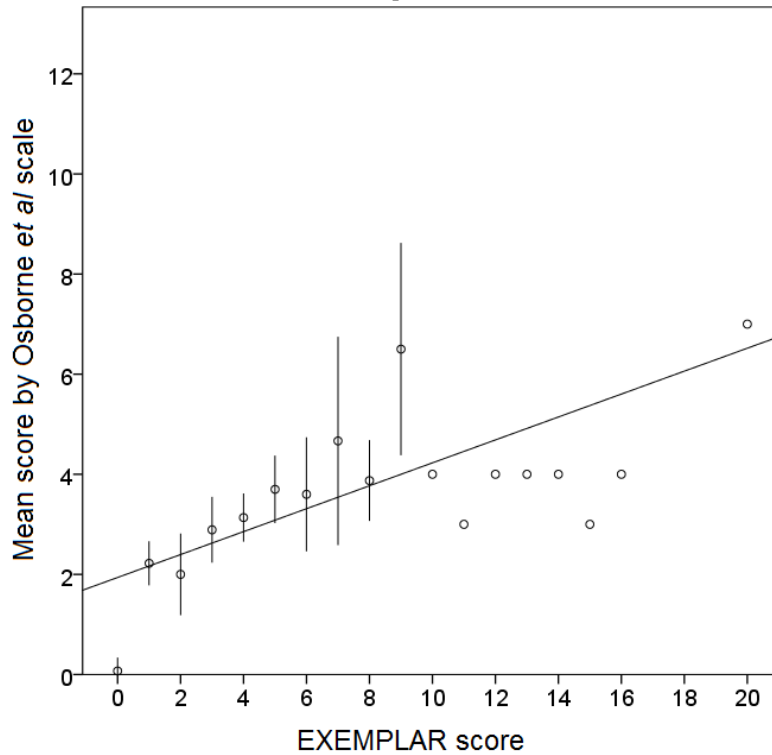
Figure 5. The mean "Osborne score" for journals with the same EXEMPLAR classification, with an overlaid best-fit regression line ($R^2 = 0.39$). Error bars represent $\pm$ 1 standard deviation of the mean for "Osborne score", for a confidence interval of 95%. No journal was classified with an EXEMPLAR score of 17, 18 or 19 points. Correlation was stronger (Pearson's $r = 0.736$), for lower scoring journals (under 10 points in the EXEMPLAR scale, corresponding to 91% of the sample).
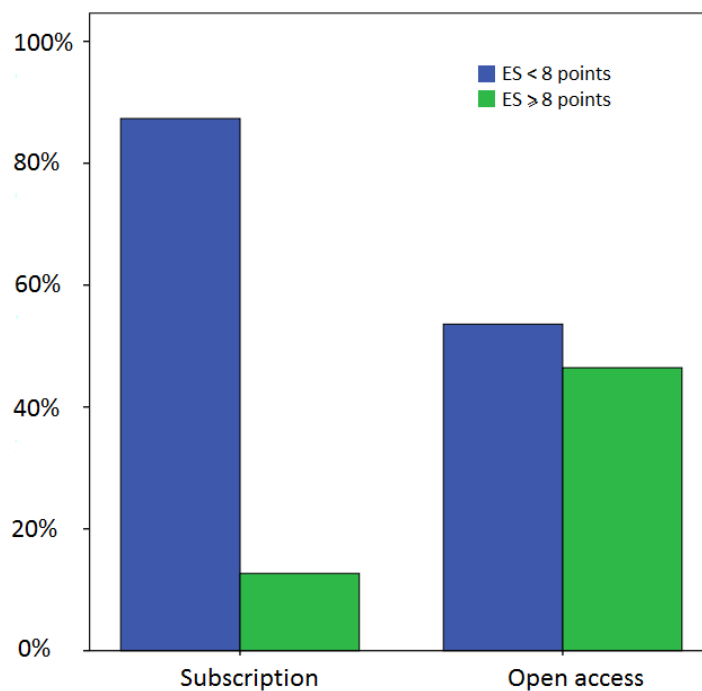

Figure 6. The proportion of open access journals with an ES $\geq$ 8 is significantly higher than that of subscription-based journals.

4. Discussion and Conclusions

Our results suggest a more positive scenario than previous reviews of journal policies [27–29], as some level of concern is patent by most journals in our sample requiring a statement on regulatory compliance. However, the little progress found regarding in-house policies on the ethical treatment of animals is worrisome.

Both the EXEMPLAR and 'Osborne *et al.*'s scoring of our sample paint the same overall scenario, and scores of the two classification schemes were, thus, not surprisingly, correlated. However, the EXEMPLAR allowed for a much higher differentiation between journals' policies; for any given group of journals with the same 'Osborne *et al.*' score, several EXEMPLAR scores were often given; for instance, the 94 journals scoring three points by Osborne *et al.*'s scale were classified under ten different EXEMPLAR scores, ranging between one and 15 points. The latter included the journal with the second-most comprehensive and stringent policy on animal use, namely *PLOS Genetics*, with an *EXEMPLAR score = 15:(5,5,5,0)*. Additionally, the top-scoring journal *PLOS One* [*EXEMPLAR score= 20*:(*5,5,5,5*)] only received six points by the 'Osborne *et al.*' classification'. These discrepancies likely result from the EXEMPLAR covering some relevant items not contemplated by Osborne *et al.*'s scale, as well as from the different weight given to some items on the overall score. Also, we consider the EXEMPLAR to be more grounded on what in the current landscape can reasonably be expected from a journal with the highest standards, whereas some of the 'Osborne *et al.*' criteria could be considered more unrealistic (e.g., expecting journals to "demand that authors 'improve on minimum standards set out in the relevant legislation'") or unnecessary (for instance, no journals demanded that "investigators and all personnel who handle and use animals are appropriately trained and qualified", a likely consequence of journals assuming that only adequately trained professionals are granted licenses to work with animals).

Of all the parameters analyzed, only model of publication showed to significantly impact the overall EXEMPLAR score, which may be a consequence of open access journals being more freely accessible and hence more open to public scrutiny. This finding apparently contradicts a previous report [29] that accessibility to articles played no influence on the ethical stance of scientific journals on animal use. The latter, however, focused only on whether journals "required demonstrating adherence to any ethical guidelines", a very superficial view of ethical standards that by the EXEMPLAR scale would be solely covered by Category A. The said study also proposed that older journals were more likely to require authors to provide a statement of regulatory compliance, a claim also challenged by our data. It has also been suggested that high impact factor journals may request less details on the animal experiments they publish [24]. At least judging from what is explicitly required by journals in their instructions for authors, we have not found any evidence that such a relationship could exist.

Most journals received one or more points for Category A, for regulatory compliance, in line with previous findings that statements of compliance with animal care guidelines and ethical approval of protocols have been on the rise and are now almost universally present in biomedical papers [30,31]. However, and considering the overall poor performance of journals in the other categories, this may imply that publishers transfer their responsibilities regarding the ethical treatment and use of animals to the institutions where the studies are carried out, or its regulators. This is particularly troublesome since the mere statement of compliance to animal care guidelines, or that a given project has been approved by regulators, do not necessarily mean that said project has gone through a thorough and balanced harm-benefit

(*i.e.*, "ethical") appraisal. Moreover, even if a project is indeed evaluated by a third party, principles and practice may vary greatly between different ethics committees (or equivalent), institutions or countries, as shown for instance by Plous and Herzog in 2001, who found an 80% disagreement between the decisions of different institutional animal care and use committee (IACUCs) over the same protocol [32].

From a regulatory point-of-view, project evaluation is often a "hybrid" between external (*i.e.*, legislation) and internal (by the scientific community) regulation of animal research, as it is usually a legal requirement, but on the other hand primarily carried out by scientists or others working in scientific institutions [33], which can lead to biases and conflicts of interests. Moreover, even if ethical appraisal is impartially carried out by knowledgeable and competent persons, this does not guarantee that the process is followed by any supervision or retrospective assessment, or that researchers going beyond the limits of their license are held accountable [34]. In fact, unacceptably severe studies (e.g., with death as an endpoint) may indeed be more prone to report any kind of regulatory compliance [30,31]. Despite these issues, since project approval does involve a third party to make a harm-benefit evaluation of the project, it was decided that it should have more weight on the overall score than the mere statement of compliance with norms and guidelines by authors, as these might be prompted by journal policy requirements, rather than having been in fact observed during the planning and execution of the experiments.

Although most journals had no explicit policies in place for the quality of reporting—covered by Category B—this category nonetheless obtained a higher mean score than the categories covering the ethical treatment of animals and the existence of criteria for the exclusion of papers. Indeed, 18% of journals were awarded the maximum score for this category, in virtually all of the cases for referring authors to the ARRIVE guidelines [23]. We opted to acknowledge journals for referring authors to reporting guidelines, even when compliance with said guidelines is not mandatory. The reasons behind our choice were, firstly, to be more realistic as regards to the current landscape, since only one journal in our sample—*PLOS Medicine*—made it mandatory to submit an ARRIVE checklist at the time of our analysis (*PLOS ONE* has, however, very recently changed its policies and now also requires submission of an ARRIVE checklist). Secondly, the actual enforcing of reporting guidelines relies on a clear understanding and communication between editors, reviewers and authors on this matter, which cannot be assessed by any analysis of the information made available in journals' websites.

It is encouraging that journals are increasingly referring authors to reporting guidelines. However, the level of adherence is still very low, in line with recent reports that nearly half of the editors in veterinary journals are unaware of any guidelines for reporting of animal experiments [35]. The upholding of reporting standards by scientific journals is of central importance to allow an informed interpretation of published results, as well as to allow replication of the experiments, a cornerstone of scientific research. Also, having sufficient information on the experimental protocol and results available is of central importance for systematic reviews and meta-analyses, which provide evidence of treatment effects to inform clinical research [24]. Journals requiring detailed information on experimental design and statistics can also help improve the quality of research itself, as researchers wanting to publish their results in journals with high reporting standards must consider best practice in the project planning phase. Furthermore, currently available checklists listing key parameters of sound animal research (e.g., power analysis, random assignment of animals to treatment groups or blinding of observers [36]) or animal welfare (such as cage enrichments, pain alleviation or humane endpoints) can help researchers verify whether these are in place when planning and executing an experiment [37]. Attention to experimental

design and adequate statistics can also allow for using only as many animals as scientifically necessary, hence following the 3Rs principle of *reduction* [36]. Additionally, improving journals' adherence to reporting guidelines can help prevent the waste of animals' lives and resources on irreproducible—on account of insufficient information in papers—or unreliable studies skewed by unaccounted biases and methodological errors [38–41].

The high number of nil-scoring journals for Category C—"Ethical treatment of animals"—is particularly worrisome, as the ethical [42], social [8] and legal [16] acceptability of animal experiments is typically grounded on a harm-benefit balance, which not only warrants animal use to be adequately justified, but also requires that experiments cause the least possible impact on animal welfare. It is also the category that focuses more on journals' proactivity in promoting the ethical treatment of animals in research, thus becoming, as proposed by Bernard Rollin, true "guardians of the gates for animal welfare" [26]. The EXEMPLAR deliberately places a greater focus on refinement in this category, from the assumption that it is more reasonable to expect that editors and reviewers are willing—and have the needed expertise—to deal with the question of *how* studies submitted for publication were carried out than to judge if these should have rather been carried out by non-animal methods in the first place.

There are, however, limitations to what can be ascertained about how animals are treated from what researchers report, but editors and reviewers with reasons to question animal welfare standards should be able to ask authors the needed information to clarify this matter and determine if indeed animal suffering was excessive or unjustified. Furthermore, there is only so much that can be asked of reviewers, who for each paper they are asked to review have to carry out a specialized and unpaid work in a short period of time [43]. Also, the reviewers' main task is usually to appraise manuscripts from a scientific point-of-view, being thus chosen for their expertise on a given field, rather than on ethics and animal welfare. These may hence lack the knowledge—and sometimes even the empathy—needed to adequately assess whether "all efforts were made to minimize suffering" (the standard, almost *cliché*, statement that can often be found in papers reporting unacceptably severe studies), or critically evaluate if the soundness and value of the study justify the cost borne by animals. Hence, editors' duty must go beyond merely requiring a statement of compliance with local regulations [28,34,44].

The overwhelming majority of journals had no explicit statement on the exclusion of manuscripts for the unethical treatment of animals, covered by Category D. The absence of a statement demanding compliance with journal policies for manuscript acceptance does not necessarily imply that editors do not enforce them, but the fact that no measurable progress in the quality of reporting has been observed in journals endorsing the ARRIVE guidelines [45] raises the question of how effective journal policies are in promoting best practice in other aspects of animal research, as well, including those measured by the EXEMPLAR scale.

The conclusions that can be drawn from the data presented here have some limitations. Firstly, our sample was retrieved from a list of papers written in English, leaving out several scientific journals publishing in other languages such as Spanish, Portuguese, Mandarin or Russian, to name a few. Also, the EXEMPLAR scale was designed to classify explicitly stated journal policies, so scores may thus reflect "good intentions" rather than actual attitudes and practice by editors and reviewers. On the other hand, there might also be cases in which the opposite is true, as editors and reviewers who are sensitive to scientific and ethical issues on animal use may have their own set of criteria and apply them accordingly, regardless of explicitly stated journal policies. There may also be an inherent bias in our

journal selection process, which aimed for high representativity of the selected fields (as the journals in the sample represent 75%, 73% and 72% of all ALS, T1D and TB papers published between 2011 and 2013, respectively) but on the other hand potentially limiting the extent to which these results can be extrapolated to the whole population of biomedical journals. There is, however, considerable diversity between the selected fields, and all major publishers were represented, as well as generalist journals, such as *PLOS ONE*, *Nature*, *PNAS* or the *EMBO journal*, among others. Furthermore, of all the journals retrieved from the original search results, only one was excluded for not publishing animal studies.

The disappointingly low scores found in our sample led us to consider the possibility that the EXEMPLAR could be setting unachievable standards. While a few high-scoring journals were found in our sample, we nevertheless analyzed a selection of five journals outside our original sample, reputed for their high standards regarding the ethical treatment of animals in research. These included two journals on laboratory animal science—*JAALAS* [EXEMPLAR =**16**:(2,5,5,4)] and *Laboratory Animals* [EXEMPLAR = **18**:(5,5,5,3)]—two others on animal welfare—*Animal Welfare* [EXEMPLAR= **15**:(4,5,3,3) and *Journal of Applied Animal Welfare* [EXEMPLAR=**12**:(2,2,5,3)]—and also MDPI's *Animals* [EXEMPLAR = **15**:(5,5,5,3). These scores attest to the construct validity of the EXEMPLAR scale (*i.e.*, the degree to which it the scale measures what it intends to be measuring), as it allowed identifying quantitatively the (subjectively perceived) high standards of this selection of journals.

Overall, our data suggest there is still much to be done as regards the attention given by biomedical journals to the scientific and ethical issues underlying animal research. The main challenge will be finding how to make journal policies promote best practice, when most journals have failed [45] to ensure adherence to reasonable [44] principles of transparency and rigorous reporting, despite the set of guidelines and checklists now available for this purpose. A checklist for journal policies on animal welfare could hence also fail to have a meaningful impact in laboratory animals' lives, because unlike reporting guidelines—that can be more broadly adopted—defining animal welfare requirements warrants a case-by-case approach, depending on the type of model used, choice of procedures and the scientific objectives. This may partially explain why compliance with regulations is a requirement for most journals, while having in-house policies in place on the ethical treatment of animals is not.

In conclusion, a re-evaluation of scientific journals' current policies, as well as an improvement in the communication between editors and reviewers to implement such policies is of the essence [44]. A discussion of these results and of the standards upheld by the EXEMPLAR may be a good starting point.

Acknowledgments

Author Contributions

NHF conceived and designed the experiments; ARM performed the experiments; ARM and NHF analyzed the data; ARM and NHF wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Franco, N.H. Animal experiments in biomedical research: A historical perspective. *Animals* **2013**, *3*, 238–273.
2. Hagelin, J.; Carlsson, H.-E.; Hau, J. An overview of surveys on how people view animal experimentation: Some factors that may influence the outcome. *Public Underst. Sci.* **2003**, *12*, 67–81.
3. Ormandy, E.H.; Schuppli, C.A.; Weary, D.M. Public attitudes toward the use of animals in research: Effects of invasiveness, genetic modification and regulation. *Anthrozoos: Multidiscip. J. Interact. People Anim.* **2013**, *26*, 165–184.
4. European Commission. Results of Questionnaire for the General Public on the Revision of Directive 86/609/EEC on the Protection of Animals Used for Experimental and Other Scientific Purposes. Available online: http://ec.europa.eu/environment/chemicals/lab_animals/pdf/results_citizens.pdf (accessed on 12 November 2014).
5. Lund, T.B.; Sørensen, T.I.A.; Olsson, I.A.S.; Hansen, A.K.; Sandøe, P. Is it acceptable to use animals to model obese humans? A critical discussion of two arguments against the use of animals in obesity research. *J. Med. Ethics* **2014**, *40*, 320–324.
6. Aldhous, P.; Coghlan, A.; Copley, J. Animal experiments—Where do you draw the line?: Let the people speak. *New Sci.* **1999**, *162*, 26–31.
7. Worcester, R.M. Science and society: What scientists and the public can learn from each other. *Proc. R. Inst.* **2001**, *71*, 97–160.
8. Lund, T.B.; Mørkbak, M.R.; Lassen, J.; Sandøe, P. Painful dilemmas: A study of the way the public's assessment of animal research balances costs to animals against human benefits. *Public Underst. Sci.* **2014**, *23*, 428–444.
9. TNS Opinion & Social. *Special Eurobarometer 340/Wave 73.1: Science and Technology*. Brussels; Directorate General Research, European Commission. **2010**
10. Pifer, L.; Shimizu, K.; Pifer, R. Public attitudes toward animal research: Some international comparisons. *Soc. Anim.* **1994**, *2*, 95–113.
11. Serpell, J.A. Factors influencing human attitudes to animals and their welfare. *Anim. Welf.* **2004**, *13*, S145–S152.
12. Holmberg, T.; Ideland, M. Secrets and lies: "Selective openness" in the apparatus of animal experimentation. *Public Underst. Sci.* **2012**, *21*, 354–368.
13. Von Roten, F.C. Public perceptions of animal experimentation across Europe. *Public Underst. Sci.* **2012**, doi: 10.1177/0963662511428045.
14. Russell, W.M.S.; Burch, R.L. *The Principles of Humane Experimental Technique*; Methuen & Co. Ltd: London, UK, 1959.
15. Festing, S.; Wilkinson, R. The ethics of animal research. *EMBO Rep.* **2007**, *8*, 526–530.

16. Louhimies, S. Eu directive 2010/63/eu: "Implementing the three rs through policy". *ALTEX Proc.* **2012**, *1*, 27–33.

17. Dixon-Woods, M.; Ashcroft, R. Regulation and the social licence for medical research. *Med. Health Care Philos.* **2008**, *11*, 381–391.

18. Varga, O.; Hansen, A.K.; Sandoe, P.; Olsson, I.A.S. Improving transparency and ethical accountability in animal studies. *Embo Rep.* **2010**, *11*, 500–503.

19. Wells, D.J. Animal welfare and the 3rs in european biomedical research. *Ann. N.Y. Acad. Sci.* **2011**, *1245*, 14–16.

20. Jennings, M.; Smith, J.A. Ethical review of animal experiments: Current practice and future directions. *ALTEX Proc.* **2012**, *1*, 275–279.

21. Forni, M. Laboratory animal science: A resource to improve the quality of science. *Vet. Res. Commun.* **2007**, *31*, 43–47.

22. Smaje, L.H.; Smith, J.A.; Combes, R.D.; Ewbank, R.; Gregory, J.A.; Jennings, M.; Moore, G.J.; Morton, D.B. Advancing refinement of laboratory animal use. *Lab. Anim.* **1998**, *32*, 137–142.

23. Kilkenny, C.; Browne, W.J.; Cuthill, I.C.; Emerson, M.; Altman, D.G. Improving bioscience research reporting: The arrive guidelines for reporting animal research. *PLoS Biol.* **2010**, *8*, e1000412.

24. Hooijmans, C.R.; Leenaars, M.; Ritskes-Hoitinga, M. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the three rs, and to make systematic reviews more feasible. *ATLA-Altern. Lab. Anim.* **2010**, 38, 167-182.

25. Basel Declaration Society. Basel declaration: A call for more trust, transparency and communication on animal research. In *Research at a Crossroads*; Basel Declaration Society: Basel, Switzerland, **2010**.

26. Rollin, B.E. Animal research, animal welfare, and the three Rs. Available online: http://jpsl.org/archives/animal-research-animal-welfare-and-three-rs/ (accessed on 3 November 2014).

27. Osborne, N.J.; Phillips, B.J.; Westwood, K. Journal editorial policies as a driver for change—Animal welfare and the 3Rs. In *New Paradigms In Laboratory Animal Science*, Proceedings of the Eleventh FELASA Symposium and the 40th Scand-LAS Symposium, Helsinki, Finland, 14–17 June 2010; FELASA: Helsinki, Finland, 2011, pp. 18–23.

28. Osborne, N.J.; Payne, D.; Newman, M.L. Journal editorial policies, animal welfare, and the 3Rs. *Am. J. Bioeth.* **2009**, *9*, 55–59.

29. Rands, S., Ethical policies on animal experiments are not compromised by whether a journal is freely accessible or charges for publication. *Animal* **2009**, *3*, 1591–1595.

30. Franco, N.H.; Olsson, I. "How sick must your mouse be?"—An analysis of the use of animal models in huntington's disease research. *ATLA* **2012**, *40*, 271–283.

31. Franco, N.H.; Correia-Neves, M.; Olsson, I.A.S. Animal welfare in studies on murine tuberculosis: Assessing progress over a 12-year period and the need for further improvement. *PLOS ONE* **2012**, *7*, e47723.

32. Plous, S.; Herzog, H. Animal research. Reliability of protocol reviews for animal research. *Science* **2001**, *293*, 608–609.

33. Varga, O.; Olsson, I.A.S. 6.2. Animal ethics committees: Different regions—Imilar tasks and difficulties?—European union. In *Ethical Aspects of Animal Research—Life in the Lab*; Gjerris, M., Olsson, I.A.S., Röcklinsberg, H., Eds.; Springer: in press.

34. Franco, N.H.; Olsson, I. Is the ethical appraisal of protocols enough to ensure best practice in animal research? *ATLA* **2013**, *41*, P5–P7.

35. Grindlay, D.J.; Dean, R.S.; Christopher, M.M.; Brennan, M.L. A survey of the awareness, knowledge, policies and views of veterinary journal editors-in-chief on reporting guidelines for publication of research. *BMC Vet. Res.* **2014**, *10*, 10.

36. Festing, M.F.; Altman, D.G. Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J.* **2002**, *43*, 244–258.

37. Hooijmans, C.R.; de Vries, R.; Leenaars, M.; Curfs, J.; Ritskes-Hoitinga, M. Improving planning, design, reporting and scientific quality of animal experiments by using the gold standard publication checklist, in addition to the arrive guidelines. *Br. J. Pharmacol.* **2011**, *162*, 1259–1260.

38. Van der Worp, H.B.; Howells, D.W.; Sena, E.S.; Porritt, M.J.; Rewell, S.; O'Collins, V.; Macleod, M.R. Can animal models of disease reliably inform human studies? *PLOS Med.* **2010**, *7*, e1000245.

39. Landis, S.C.; Amara, S.G.; Asadullah, K.; Austin, C.P.; Blumenstein, R.; Bradley, E.W.; Crystal, R.G.; Darnell, R.B.; Ferrante, R.J.; Fillit, H.; *et al.* A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* **2012**, *490*, 187–191.

40. Kilkenny, C.; Parsons, N.; Kadyszewski, E.; Festing, M.F.; Cuthill, I.C.; Fry, D.; Hutton, J.; Altman, D.G. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLOS ONE* **2009**, *4*, e7824.

41. Eisen, J.A.; Ganley, E.; MacCallum, C.J. Open science and reporting animal studies: Who's accountable? *PLOS Biol.* **2014**, *12*, e1001757.

42. Olsson, I.A.S.; Robinson, P.; Sandøe, P. Ethics of animal research. *Handb. Lab. Anim. Sci.* **2011**, *1*, 21–35.

43. Hirst, A.; Altman, D.G. Are peer reviewers encouraged to use reporting guidelines? A survey of 116 health research journals. *PLOS ONE* **2012**, *7*, e35621.

44. Galley, H.F. Mice, men, and medicine. *Br. J. Anaesth.* **2010**, *105*, 396–400.

45. Baker, D.; Lidster, K.; Sottomayor, A.; Amor, S. Two years later: Journals are not yet enforcing the arrive guidelines on reporting standards for pre-clinical animal studies. *PLOS Biol.* **2014**, *12*, e1001756.