



## Ensuring reproducibility and ethics in animal experiments reporting in Korea using the ARRIVE guideline

Mi-Hyun Nam<sup>1,2</sup>, Myung-Sun Chun<sup>3</sup>, Je-Kyung Seong<sup>1,2,\*</sup>, Hoon-Gi Kim<sup>4,\*</sup>

<sup>1</sup>Laboratory of Developmental Biology and Genomics, College of Veterinary Medicine, Research Institute of Veterinary Science, and BK21 Program for Veterinary Science, Seoul National University, Seoul, Korea

<sup>2</sup>Korea Mouse Phenotyping Center (KMPC), Seoul National University, Seoul, Korea

<sup>3</sup>Research Institute of Veterinary Science, College of Veterinary Medicine, Seoul National University, Seoul, Korea

<sup>4</sup>College of Liberal Arts, Hongik University, Seoul, Korea

The aim of this study is to evaluate the reporting quality of animal experiments in Korea using the Animals in Research: Reporting *In Vivo* Experiments (ARRIVE) guideline developed in 2010 to overcome the reproducibility problem and to encourage compliance with replacement, refinement and reduction of animals in research (3R's principle). We reviewed 50 papers published by a Korean research group from 2013 to 2016 and scored the conformity with the 20-items ARRIVE guideline. The median conformity score was 39.50%. For more precise evaluation, the 20 items were subdivided into 57 sub-items. Among the sub-items, status of experimental animals, housing and husbandry were described under the average level. Microenvironment sub-items, such as enrichment, bedding material, cage type, number of companions, scored under 10%. Although statistical methods used for the studies were given in most publications (84%), sample size calculation and statistical assumption were rarely described. Most publications mentioned the IACUC approval, but only 8% mentioned welfare-related assessments and interventions, and only 4% mentioned any implications of experimental methods or findings for 3R. We may recommend the revision of the present IACUC proposal to collect more detailed information and improving educational program for animal researchers according to the ARRIVE guideline.

**Keywords:** Animal experiments, ARRIVE guideline, reproducibility crisis, 3R's principle

Received 16 November 2017; Revised version received 30 January 2018; Accepted 30 January 2018

In a recent survey of 1,576 researchers, 52% of the respondents perceived the crisis of reproducibility and the increasing inability to reproduce another scientist's experiments. Over 60% of scientists in biology and medicine reported not being able to reproduce results by other researchers [1]. In the past decade, academic journals have noted the scientific and ethical implications for the entire research process due to the failure to describe research methods and to report results comprehensibly and accurately [2]. The main challenges were found in statistical methods including randomization and blinding,

and in describing status of experimental animals. For example, in a survey of 271 publications on reporting research methods, only 59% of the studies described the number and characteristics of the used animals, and 30% of the studies did not present statistical information. Less than 20% of the studied publications mentioned randomization or blinding for reducing bias in the experiments [3].

The 3R's principle (replacement, refinement and reduction of animals in research) in the perspective of quality improvement of science has been recognized

\*Corresponding authors: Je-Kyung Seong, Room 514, Bldg 81, College of Veterinary Medicine, Seoul National University, Gwanak-ro 1, Gwanak-gu, Seoul, 08826, Korea

Tel: +82-2-880-1259; Fax: +82-2-875-8395; E-mail: [snumouse@snu.ac.kr](mailto:snumouse@snu.ac.kr)

Hoon-Gi Kim, Room 421, College of Liberal Arts, Hongik University, Wausan-ro 94, Mapo-gu, Seoul, 04066, Korea

Tel: +82-2-320-3362; Fax: +82-2-320-3299; E-mail: [wolfkim8@gmail.com](mailto:wolfkim8@gmail.com)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

because the treatment of experimental animals in the most humane way is a prerequisite for a successful animal experiment [4-6]. It is not limited to the implementation phase but whole process of scientific investigation from formulation to communication and replication. More recently, the 3R's principle has become acknowledged as an integral part of conducting high quality bioscience and a means of addressing issues of poor reproducibility and high rates of attrition in drug development [7].

The Animals in Research: Reporting *In Vivo* Experiments (ARRIVE) guideline was developed in 2010 to improve the design, analysis, and reporting of research using animals by the National Centre for the Replacement, Refinement & Reduction of Animals in Research (NC3Rs). The ARRIVE guideline provides a 20-item checklist for a minimum description of animals included in a scientific paper. Some items are study design, experimental procedure, experimental animals, housing and husbandry, and sample size [2].

Since 2010, leading biomedical journals have adopted the ARRIVE guideline for reporting research methods and results to ensure quality of the published scientific papers. In the intervening 8 years, the reporting quality in these science publications has not appreciably changed. Yet, the guideline has proven successful in motivating the scientific community to recognize the problems related to animal research and strive to improve the situation. For example, an evaluation of 83 publications assessing new compounds for Chagas disease in animal models showed slight improvement in reporting information on animals, macro- and microenvironment in spite of significant lack of the reporting accuracy [6]. Also, a review of 396 scientific studies involving animals in China showed that the publications after 2010 had a significantly larger ARRIVE value than in 2010 [8].

Genetically engineered mice (GEM) have become a major research tool to identify the function of human genes [9] and the mechanism of human disease [10]. The mouse phenotyping, which is the work of identifying the function of the mouse gene, is proceeding with international cooperation under the leadership of International mouse phenotyping consortium (IMPC). IMPC requires member organizations to report the phenotypic analysis results in detail, and to ensure reproducibility of the experiment by sharing them.

The present study examined 50 scientific papers published by the Korea Mouse Phenotyping Center

(KMPC), an IMPC member and one of the largest mouse research teams in Korea, to examine their conformity with the ARRIVE guideline. This evaluation focuses on accurate description of experimental animals and statistical methods from the reproducibility perspective and ethical consideration for experimental animals from the 3R's perspective. Revealing the weaknesses and strengths in reporting animal experimentation in Korea, and the need to enhance research quality by adopting *in vivo* experiments reporting guidelines in our science community can only be beneficial.

## Materials and Methods

### Selection of papers

Titles and abstracts of 88 papers published in English in academic journals between 2013 and 2016 by KMPC members were screened. Papers using animals were selected by the authors including an experimental animal veterinarian. The full texts with supplementary materials were reviewed in cases of difficulty in selecting based on titles and abstracts. The potentially proper papers were finally selected by excluding ones beyond the scope of this study. For example, papers on *in vitro* study and reviews were excluded.

### Evaluation of papers

The ARRIVE guideline consists of 20 checklist items and 39 sub-items. To provide more detailed information on the reporting situation, we added 18 sub-items to the checklist by subdividing items related to experimental animals (item 8), housing and husbandry (item 9), and study design about minimization of subjective bias (item 6).

All the 57 sub-items were assessed as "YES (described in the study)" or "NO (not described in the study)". If a sub-item could not be assessed due to the research type (for example, "quality of water for fish" in the case of mouse experiments), it was indicated as "NONE".

### Data analysis

The data were summarized using Microsoft Excel. Evaluation scores are expressed as the ratio of absolute numbers and percentiles in the 20 items and 57 sub-items. The number of "NONE" cases was excluded from the calculations. All the scores about the papers of animal experiments were presented at Figure 1~5 and Table 1.

## Results

### Selected papers

Of the 88 KMPC papers, 50 papers were selected for evaluation concerning their application of the ARRIVE guideline. Among them, 41 were *in vivo* studies that compared test groups with control groups as well as analyzing specific phenotypes, with the remainder being done *ex vivo* conditions. Thirty seven papers were excluded in the screening of titles and abstracts, with one more excluded after the assessment of the full texts.

### Score distribution by items and sub-items

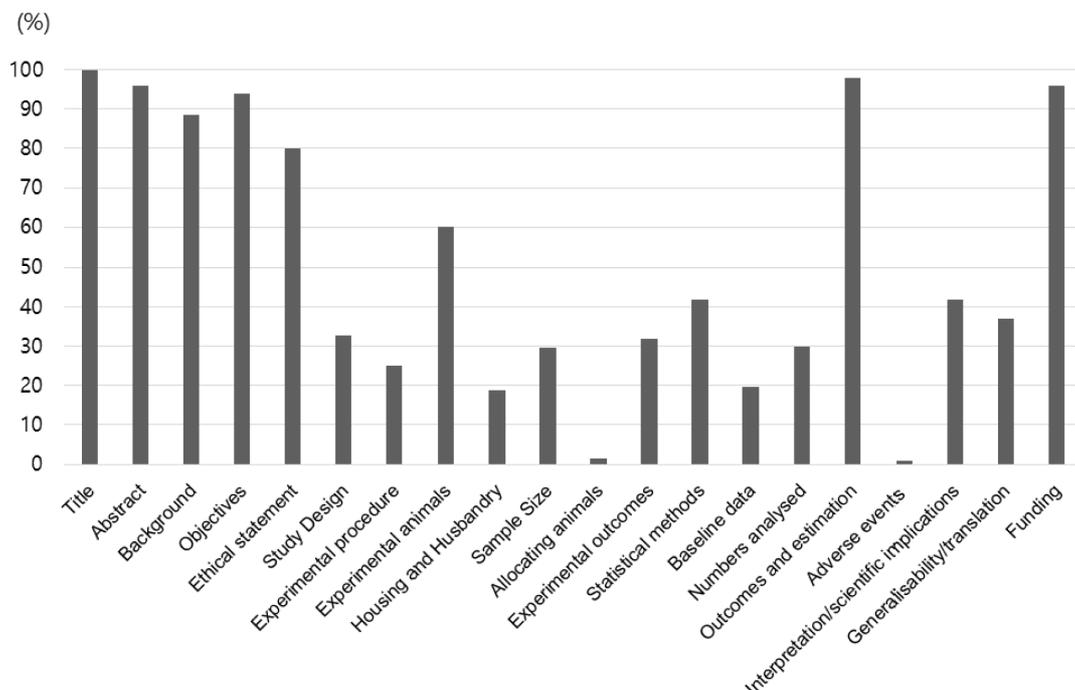
The scores for 20 ARRIVE items are shown in Figure 1. The median percentile score was 39.50% (average 51.17%). The five most frequently reported items were title, abstract, objectives, outcomes and estimation, and funding. In contrast, five items that were least frequently reported were experimental procedural items of housing and husbandry, allocating animals, baseline data, and adverse events. The scores for the 57 sub-items are presented in Table 1. The median score was 30.00% (average 39.94%). Sub-items scored over 75% were 1, 2, 3-a, 4, 5, 7-a, 8-a1, 8-a2, 8-a4, 8-b1, 8-b2, 13-a, 16, 18-a, and 20. These sub-items mainly correspond to

information on title, abstract, background, objectives, experimental animals, outcome and estimation, and funding. In contrast, sub-items scored under 25% were 6-b1, 6-b2, 7-b, 7-c, 7-d, 8-a5, 8-b3, 9-a2, 9-a3, 9-a4, 9-b3, 9-c, 10-b, 11-a, 14, 17-a, 18-b, and 18-c. Sub-items 8-b4, 9-b4, 9-b7, 11-b, 13-c and 17-b were not described. These sub-items are chiefly related to experimental procedure, housing and husbandry, allocating animals, and adverse events.

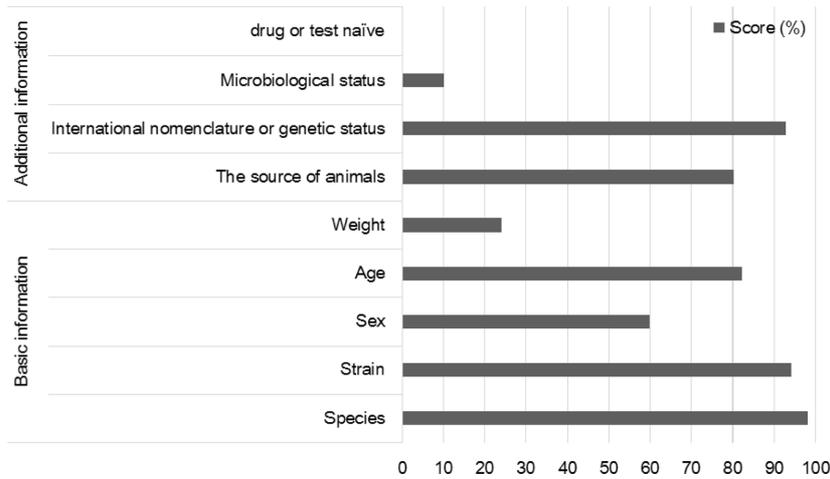
### Reporting of experimental animals and housing and husbandry

The reporting scores on the experimental animals are shown in Figure 2, and are categorized as basic and additional information. Basic information of experimental animals, such as strain, sex, and age, were described frequently. Animal weight was not, being mentioned in only 25% of papers. Microbiological status and information on drug or test history of animal used were described in fewer than 10% of the papers.

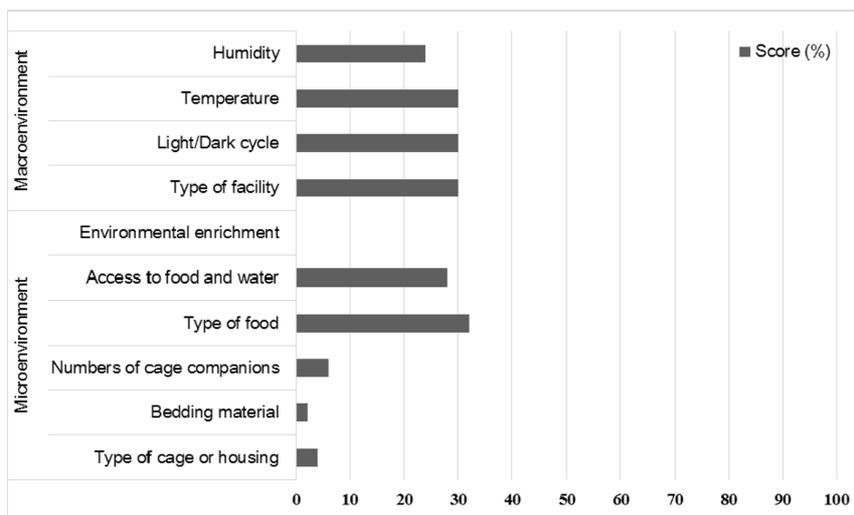
Information on microenvironment and macroenvironment were described in the sub-items of housing and husbandry (Figure 3). Microenvironment refers to the immediate physical environment surrounding the animal, such as type of cage, bedding material, numbers of cage companion,



**Figure 1.** Scores (%) of the 20 items of the ARRIVE guideline. Items of title, abstract, objectives, outcomes and estimation, and funding were the most frequently reported (over 80%). The items of experimental procedure, housing and husbandry, and baseline data were the least frequently reported (under 30%). Allocating animals and adverse events were items reported under 5%.



**Figure 2.** Scores (%) of the sub-items on experimental animals. Basic information on experimental animals (age, sex, strain, and species) was reported in over 60% of the reviewed publications. Although the additional information on the animals, such as the source of animals and genetic status, were mentioned in over 80% of the publications, information on microbiological status and drug test were rarely stated.

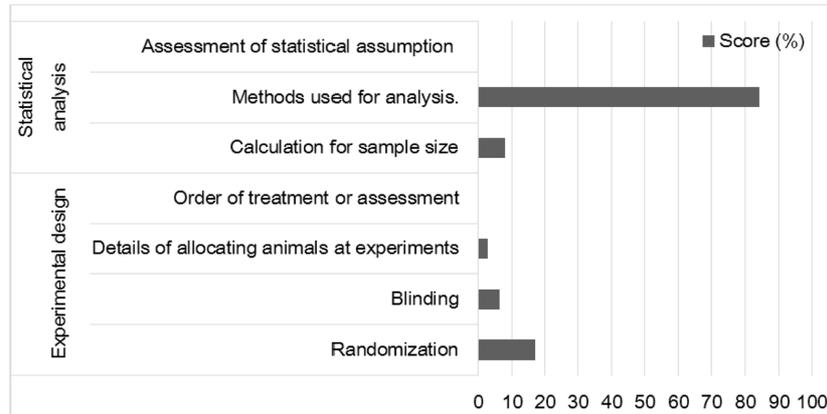


**Figure 3.** Scores (%) of sub-items on housing and husbandry. Macroenvironment (humidity, temperature, light/dark cycle and type of facility) were described in less than 30% of the reviewed publications. Little information on microenvironment (type of cage, bedding material, numbers of cage companion, type of food and access to food/water) was reported. No publication reported environmental enrichment.

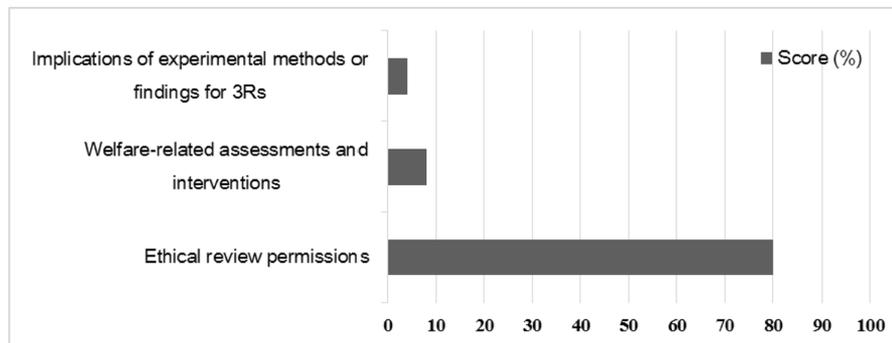
type of food, access to food/water and environmental enrichment. Macroenvironment parameters include the physical environment of the secondary enclosure, such as type of facility, light/dark cycle, temperature, and humidity [11]. Information about housing and husbandry were described in fewer than 30% of the reviewed publications. In particular, the sub-items of micro-environment, such as enrichment, bedding material, cage type, number of companions, scored under 10%. No paper described environmental enrichment.

**Reporting of experimental design and statistical analysis**

Most papers did not report randomization in animal selection or blinding in outcome measurements to minimize the effects of subjective bias (Figure 4). Methods about allocating animals to experimental groups and order for treatment or assessment were rarely mentioned. Score of the statistical methods used for each analysis was relatively high at 84%, while only 8% of the papers described the sample size calculation. No



**Figure 4.** Scores (%) of sub-items on experimental design and statistical analysis. Except statistical methods used for analysis, the sub-items on experimental design and statistical analysis were rarely found in the reviewed publications.



**Figure 5.** Scores (%) of sub-items on ethical statement and 3R's principle. Because the IACUC permission is legal prerequisite for animal experiment 80% of the reviewed publications have ethical statement about protocol review and the IACUC approval.

study reported about assessment of statistical assumption.

### Reporting of ethical statement and 3R's principle

The results of assessment about laboratory animal ethics are shown in Figure 5. Reporting for the ethical statement about protocol review and approval by Institutional Animal Care and Use Committee (IACUC) was scored 80%. Of these, 34% described the approval number, while the remaining 46% did not. 16% of the papers just mentioned like “follow affiliation’s guide” and 4% did not include a specific ethical statement. Only 8% of papers mentioned welfare-related assessments and interventions. Only 4% of paper mentioned any implications of experimental methods or 3R findings of the study.

## Discussion

The ARRIVE guideline was developed to ensure reproducibility of animal experiments and to avoid

unnecessary animal use. It also aid in more transparent, comprehensive, and logical reporting of research findings. A recent phenomenon described as a ‘reproducibility crisis’ in science can be effectively overcome by following this guideline at design, conduct, and analysis [12-14].

Of the reviewed publications, only 30-40% included the required ARRIVE information about their experiments. Considering that the publications were produced in an animal experiment-specialized group, the score represents the maximum level of the conformity with the standards.

These scores, however, do not directly correspond to the reproducible level of papers. The ARRIVE guideline is a ‘reporting’ guideline aimed to improve reproducibility and ethical consideration in published papers. Reported experimental steps could be omitted because the peer reviewers thought the description was unnecessary. Some sub-items of the ARRIVE guideline could not be applied because of the characteristics of the experiments. However, it is clear that there is a need for improvement

**Table 1.** Scoring of the 57 sub-items of the ARRIVE guideline

Item	Sub-items*	Ratio	Score (%)
TITLE	1 Provide as accurate and concise a description of the content of the article as possible.	50/50	100.00
ABSTRACT	2 Provide an accurate summary of the background, research objectives, including details of the species or strain of animal used, key methods, principal findings and conclusions of the study.	48/50	96.00
INTRODUCTION			
Background	3-a Include sufficient scientific background (including relevant references to previous work) to understand the motivation and context for the study, and explain the experimental approach and rationale.	49/50	98.00
	3-b Explain how and why the animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology	36/49	73.47
Objectives	4 Clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested.	47/50	94.00
METHODS			
Ethical statement	5 Indicate the nature of the ethical review permissions, relevant licenses (e.g. Animal [Scientific Procedures] Act 1986), and national or institutional guidelines for the care and use of animals, that cover the research.	40/50	80.00
Study Design	6-a For each experiment, give brief details of the study design including The number of experimental and control groups.	35/47	74.47
	6-b1 Any steps taken to minimize the effects of subjective bias when allocating animals to treatment (e.g. randomization procedure)	6/35	17.14
	6-b2 Any steps taken to minimize the effects of subjective when assessing results (e.g. if done, describe who was blinded and when)	3/47	6.38
	6-c The experimental unit (e.g. a single animal, group or cage of animals).	23/47	48.94
	6-d A time-line diagram or flow chart can be useful to illustrate how complex study designs were carried out	8/47	17.02
Experimental procedure	7-a For each experiment and each experimental group, including controls, provide precise details of all procedures carried out. How (e.g. drug formulation and dose, site and route of administration, anesthesia and analgesia used [including monitoring], surgical procedure, method of euthanasia). Provide details of any specialist equipment used, including supplier(s).	40/48	83.33
	7-b b. When (e.g. time of day).	1/48	2.08
	7-c c. Where (e.g. home cage, laboratory, water maze).	6/48	12.50
	7-d d. Why (e.g. rationale for choice of specific anesthetic, route of administration, drug dose used).	1/48	2.08
Experimental animals	Provide details of the animals used		
	8-a1 Species	49/50	98.00
	8-a2 Strain	47/50	94.00
	8-a3 Sex	30/50	60.00
	8-a4 developmental stage (e.g. mean or median age plus age range)	41/50	82.00
	8-a5 weight (e.g. mean or median weight plus weight range)	12/50	24.00
	Provide further relevant information		
	8-b1 the source of animals	40/50	80.00
	8-b2 international strain nomenclature / genetic modification status (e.g. knock-out or transgenic) / genotype	38/41	92.68
8-b3 health/immune status	5/50	10.00	
8-b4 drug or test naïve, previous procedures	0/48	0.00	

of the accurate reporting in micro- and macro-environment for animals, study design, and the statistics and ethical consideration in research using animals.

First, the sub-items about the status of animals and

environment, which our team mainly subdivided from the original guideline, need to be reported more accurately. Animal status sub-items, such as health status, weight, and sex, and environment sub-items including bedding

Table 1. Continued

Item	Sub-items*	Ratio	Score (%)
	Provide details of:		
	9-a1 type of facility e.g. specific pathogen free [SPF]	20/50	40.00
	9-a2 type of cage or housing	3/50	6.00
	9-a3 bedding material	1/50	2.00
	9-a4 number of cage companions	3/50	6.00
	9-a5 tank shape and material etc. for fish	1/3	33.33
	Husbandry conditions		
Housing and Husbandry	9-b1 light/dark cycle	15/50	30.00
	9-b2 Temperature	15/50	30.00
	9-b3 Humidity	12/50	24.00
	9-b4 quality of water for fish	0/3	0.00
	9-b5 type of food	17/50	34.00
	9-b6 access to food and water	15/50	30.00
	9-b7 environmental enrichment	0/50	0.00
	9-c Welfare-related assessments and interventions that were carried out prior to, during, or after the experiment.	4/50	8.00
Sample Size	10-a Specify the total number of animals used in each experiment, and the number of animals in each experimental group.	23/50	46.00
	10-b Explain how the number of animals was arrived at. Provide details of any sample size calculation used.	4/50	8.00
	10-c Indicate the number of independent replications of each experiment, if relevant.	17/49	34.69
Allocating animals to experimental groups	11-a Give full details of how animals were allocated to experimental groups, including randomization or matching if done.	1/36	2.78
	11-b b. Describe the order in which the animals in the different experimental groups were treated and assessed.	0/48	0.00
Experimental outcomes	12 Clearly define the primary and secondary experimental outcomes assessed (e.g. cell death, molecular markers, and behavioral changes).	16/50	32.00
Statistical methods	13-a Provide details of the statistical methods used for each analysis.	42/50	84.00
	13-b Specify the unit of analysis for each dataset (e.g. single animal, group of animals, single neuron).	21/50	42.00
	13-c Describe any methods used to assess whether the data met the assumptions of the statistical approach.	0/50	0.00
<b>RESULTS</b>			
Baseline data	14 For each experimental group, report relevant characteristics and health status of animals (e.g. weight, microbiological status, and drug or test naïve) prior to treatment or testing (this information can often be tabulated).	9/46	19.57
Numbers analyzed	15-a Report the number of animals in each group included in each analysis. Report absolute numbers (e.g. 10/20, not 50% <sup>2</sup> ).	24/49	48.98
	15-b If any animals or data were not included in the analysis, explain why	5/47	10.64
Outcomes and estimation	16 Report the results for each analysis carried out, with a measure of precision (e.g. standard error or confidence interval).	49/50	98.00
Adverse events	17-a Give details of all important adverse events in each experimental group.	1/50	2.00
	17-b Describe any modifications to the experimental protocols made to reduce adverse events	0/50	0.00
<b>DISCUSSION</b>			
Interpretation/scientific implications	18-a Interpret the results, taking into account the study objectives and hypotheses, current theory and other relevant studies in the literature.	49/50	98.00
	18-b Comment on the study limitations including any potential sources of bias, any limitations of the animal model, and the imprecision associated with the results.	12/50	24.00
	18-c Describe any implications of your experimental methods or findings for the replacement, refinement or reduction (the 3Rs) of the use of animals in research.	2/50	4.00
Generalizability/translation	19 Comment on whether, and how, the findings of this study are likely to translate to other species or systems, including any relevance to human biology.	18/49	36.73
Funding	20 List all funding sources (including grant number) and the role of the funder(s) in the study.	48/50	96.00

\*Items about study design (item 6), experimental animals (item 8) and housing and husbandry (item 9) were subdivided from the original checklist.

material, type of cage, and numbers of companions, are important information needed to reproduce the experimental environment. Especially, microenvironmental conditions can directly affect physiologic processes and behavior and may alter disease susceptibility [11]. In some cases of describing health status or type of facility, the term SPF (specific pathogen-free) was simply used without clarifying the specific microorganisms or whether the conditions met the standards recommended by the relevant academic society or association. To improve maximize information about environment, researchers should be familiar with the animal care program of laboratory animal facility.

The second noteworthy finding was the low scores of study design and statistics. The problem of inadequate reporting of statistics has frequently been pointed out as a major reason for the reproducibility crisis. For example, one study described that less than 10% of animal studies in *Nature* or *PLoS* journals reported randomization, and around 20% of them mentioned blinding in past surveys [15]. Also, another study found that sample size was not calculated in any of paper concerning Chagas disease that used animals [5], nor did a review article addressing the conformity with ARRIVE in China [8]. Because the importance of using correct statistical treatment methods has been increasingly emphasized in animal experiments as a way to overcome the reproducibility crisis, researchers need to report in detail the procedure for allocating animals, the number of animals before and during experiments, and the reason for adopting sample size and specific statistical methods. Especially, the description of how the sample size is reached is related to refinement of 3Rs. Therefore it can be considered as essential item that should be reported in animal using research.

Third, ethical consideration for experiments needs to be more clearly indicated in publications. According to the Animal Protection Act and the Laboratory Animals Act in Korea, all the animal experiments should be approved by the relevant IACUC. Presently, 80% of the reviewed papers mentioned the approval of IACUC or supplied the permission number. However, more active statements of the researcher's intention and active steps to maintain animal welfare during the experiments and acknowledgement of adherence to the 3R's principle are required because public concern about health and husbandry conditions is growing. Welfare-related assessments and interventions that were carried out prior

to, during, or after the experiment can be considered as post-approval monitoring (PAM) that is required by laws, regulations, and policies. PAM helps ensure the well-being of the animals and may also provide opportunities to refine research procedures [11]. For the improvement of the reporting accuracy the IACUC guidelines at the step of the study design should require the consideration of the ARRIVE guideline items. We recommend the revision of existing IACUC proposal forms with reference to the sub-items of the ARRIVE guideline to help researchers report their experiments to a standard that matches the international level. Also, an educational program on the ARRIVE guidelines would help researchers understand the reproducibility crisis and the importance of the accurate reporting of research using animals. Better understanding of the context of ARRIVE will encourage researchers to apply the guideline to improve the quality of their future studies.

## Acknowledgments

This research was supported by Korea Mouse Phenotyping Project (2014M3A9D5A01074636) of the Ministry of Science, ICT and Future Planning through the National Research Foundation. The authors thank to Hanah Sung and Misun Mang for their help in data collection.

**Conflict of interests** The authors declare that there is no financial conflict of interests to publish these results.

## References

1. Baker M. Is there a reproducibility crisis? *Nature* 2016; 533(7604): 452-454.
2. Kilkeny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010; 8(6): e1000412.
3. Kilkeny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J, Altman DG. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS one* 2009; 4(11): e7824.
4. Burden N, Chapman K, Sewell F, Robinson V. Pioneering better science through the 3Rs: an introduction to the national centre for the replacement, refinement, and reduction of animals in research (NC3Rs). *J Am Assoc Lab Anim Sci* 2015; 54(2): 198-208.
5. Gulin JE, Rocco DM, Garcia-Bournissen F. Quality of reporting and adherence to ARRIVE guidelines in animal studies for Chagas disease preclinical drug research: a systematic review. *PLoS Negl Trop Dis* 2015; 9(11): e0004194.
6. Tannenbaum J, Bennett BT. Russell and Burch's 3Rs then and now: the need for clarity in definition and purpose. *J Am Assoc Lab Anim Sci* 2015; 54(2): 120-132.
7. Graham ML, Prescott MJ. The multifactorial role of the 3Rs in

- shifting the harm-benefit analysis in animal models of disease. *Eur J Pharmacol* 2015; 759: 19-29.
8. Liu Y, Zhao X, Mai Y, Li X, Wang J, Chen L, Mu J, Jin G, Gou H, Sun W, Feng Y. Adherence to ARRIVE guidelines in Chinese journal reports on neoplasms in animals. *PloS one* 2016; 11(5): e0154657.
  9. Kim IY, Shin JH, Seong JK. Mouse phenogenomics, toolbox for functional annotation of human genome. *BMB Rep* 2010; 43(2): 79-90.
  10. Pratis S, Jurjus A. Spontaneous and transgenic rodent models of inflammatory bowel disease. *Lab Anim Res* 2015;31(2):47-68.
  11. National Research Council. Guide for the care and use of laboratory animals: National Academies Press; 2010. Available from: <http://www.nap.edu/catalog/12910.html>.
  12. Bryant FB. Enhancing predictive accuracy and reproducibility in clinical evaluation research: Commentary on the special section of the Journal of Evaluation in Clinical Practice. *J Eval Clin Pract* 2016; 22(6): 829-834.
  13. Hossenfelder S. Science needs reason to be trusted. *Nature Physics* 2017; 13(4): 316-317.
  14. Mullane K, Williams M. Enhancing reproducibility: failures from reproducibility initiatives underline core challenges. *Biochem Pharmacol* 2017; 138: 7-18.
  15. 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(1): e1001756.