



庞万勇, 博士, 中国实验动物学会认证的实验动物高级医师, 美国实验动物医学会认证的实验动物专科兽医师 (DACLAM), 以及中国兽医病理学家分会认证的兽医病理师。中国实验动物学会常务理事、实验动物医师工作委员会主任委员、实验动物福利伦理专业委员会秘书长, CNAS实验动物专业委员会委员, 中华医学会医师培训工程——中华医学人文培训工程专家委员会委员, 国际AAALAC认可委员会委员, 世界兽医协会 (WVA) 理事 (代表实验动物专科兽医师)。本科和硕士毕业于中国农业大学动物医学院, 博士毕业于爱尔兰国立都柏林大学兽医学院, 曾在丹麦哥本哈根大学从事实验动物医学博士后研究。现任赛诺菲公司全球研发中心转化体内模型研究平台总监, 负责该公司在全球范围内动物实验外包和合作业务的动物福利合规事务, 并协调该公司在中国的临床申报项目中的临床前部分。

发表文章约36篇, 其中SCI文章约20篇。

王 剑, 兽医学博士, 中国执业兽医师。毕业于华南农业大学兽医学院临床内科学, 主攻动物营养代谢病与中毒病。先后于华南农业大学畜牧学、上海市农业科学院生物学博士后流动工作站从事动物采食调控、动物病毒学等研究工作。现任上海交通大学医学院实验动物科学部实验师, 主要从事小鼠的转基因模型构建、繁育和病原微生物检测等相关工作。同时兼任上海市畜牧兽医协会动物公共卫生分会理事、中国兽医协会动物诊疗分会会员。发表国内外学术论文9篇, 参编《现代宠物医生手册》临床专著1部。



## 《动物研究:体内实验报告》即ARRIVE 2.0指南的解释和阐述(一)

王 剑<sup>1</sup>, 卢 今<sup>1</sup>, 马政文<sup>1</sup>, 陈国元<sup>2</sup>, 卢 晓<sup>3</sup>, 白 玉<sup>4</sup>, 刘晓宇<sup>5</sup>, 卢选成<sup>6</sup>, 高 静<sup>7</sup>, 李 垚<sup>1</sup>, 庞万勇<sup>8</sup>

(1. 上海交通大学医学院实验动物科学部, 上海200025; 2. 中国科学院分子细胞科学卓越创新中心动物实验技术平台, 上海200031; 3. 迪哲医药股份有限公司, 上海201203; 4. 北京诺和诺德医药科技有限公司, 北京102206; 5. 中国疾病预防控制中心职业卫生与中毒控制所, 北京100050; 6. 中国疾病预防控制中心实验动物中心, 北京102206; 7. 上海交通大学医学院临床研究中心, 上海200025; 8. 赛诺菲公司全球研发中心转化体内模型研究平台, 北京100022)

**[摘要]** 提高生物医学研究结果的可重复性是一项重大挑战, 研究人员透明且准确地报告其研究过程有利于读者对该研究结果的可靠性进行评估, 进而重复该实验或在该成果的基础上进一步探索。ARRIVE 2.0指南是2019年英国国家3Rs中心 (NC3Rs) 组织发布的一份适用于任何与活体动物研究报告相关的指导性清单, 用以提高动物体内实验设计、实验实施和实验报告的规范性, 以及动物实验结果的可靠性、可重复性和临床转化率。ARRIVE 2.0指南的使用不仅可以丰富动物实验研究细节, 确保动物实验结果信息被充分评估和利用, 还可以使读者准确且清晰地了解作者所表述的内容, 促进基础研究评审过程的透明化和完整性。目前, ARRIVE 2.0指南已经被国际生物医学期刊广泛采纳。本文是在国际期刊遵循ARRIVE 2.0指南的最佳实践基础上, 对2020年发表于*PLoS Biology*期刊上的ARRIVE 2.0指南完整解读版 (原文请见<https://arriveguidelines.org>) 进行中文编译 (第一部分包括前言和“关键10条”里的“研究设计”“样本量”“纳入和排除标准”), 以期促进国内研究人员充分理解并使用ARRIVE 2.0指南, 提高实验动物研究及报告的规范性, 助推我国实验动物科技与比较医学研究的高质量发展。

**[关键词]** 动物实验; ARRIVE 2.0指南; ARRIVE关键10条; 研究设计; 样本量; 纳入和排除标准

**[中图分类号]** R-332; Q95-33 **[文献标志码]** C **[文章编号]** 1674-5817(2023)02-0213-12

### Explanation and Elaboration for the ARRIVE Guidelines 2.0— Reporting Animal Research and In Vivo Experiments ( I )



**[第一作者]** 王 剑(1986—), 男, 兽医学博士, 实验师, 研究方向: 实验动物医学。E-mail: wangjian3976@shsmu.edu.cn

**[通信作者]** 庞万勇(1974—), 男, 兽医学博士, 研究方向: 实验动物医学。E-mail: panglyong@outlook.com。ORCID: 0000-0002-0724-2016

WANG Jian<sup>1</sup>, LU Jin<sup>1</sup>, MA Zhengwen<sup>1</sup>, CHEN Guoyuan<sup>2</sup>, LU Xiao<sup>3</sup>, BAI Yu<sup>4</sup>, LIU Xiaoyu<sup>5</sup>, LU Xuancheng<sup>6</sup>, GAO Jing<sup>7</sup>, LI Yao<sup>1</sup>, Pang Wanyong<sup>8</sup>

(1. Department of Laboratory Animal Science, School of Medicine, Shanghai Jiao Tong University, Shanghai 200025, China; 2. Animal Core Facility, Center for Excellence in Molecular Cell Science, Chinese Academy of Science, Shanghai 200031, China; 3. Dical Pharmaceuticals Co., Ltd., Shanghai 201203, China; 4. Novo Nordisk Research Centre China, Beijing 102206, China; 5. National Institute for Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing 100050, China; 6. Laboratory Animal Center, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 7. Clinical Research Institute, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; 8. Translational in Vivo Model Research Platform, Sanofi G&D, Beijing 100022, China)

Correspondence to: PANG Wanyong (ORCID: 0000-0002-0724-2016), E-mail: pang1yong@outlook.com

**[ABSTRACT]** Improving the reproducibility of biomedical research results is a major challenge. Researchers reporting their research process transparently and accurately can help readers evaluate the reliability of the research results and further explore the experiment by repeating it or building upon its findings. The ARRIVE 2.0 guidelines, released in 2019 by the UK National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs), provide a checklist applicable to any *in vivo* animal research report. These guidelines aim to improve the standardization of experimental design, implementation, and reporting, as well as the reliability, repeatability, and clinical translatability of animal experimental results. The use of ARRIVE 2.0 guidelines not only enriches the details of animal experimental research reports, ensuring that information on animal experimental results is fully evaluated and utilized, but also enables readers to understand the content expressed by the author accurately and clearly, promoting the transparency and integrity of the fundamental research review process. At present, the ARRIVE 2.0 guidelines have been widely adopted by international biomedical journals. This article is a Chinese translation based on the best practices of international journals following the ARRIVE 2.0 guidelines in international journals, specifically for the complete interpretation of the ARRIVE 2.0 guidelines published in the *PLoS Biology* journal in 2020 (original text can be found at <https://arriveguidelines.org>). The first part of the article includes the preface and the "Key 10" section, which covers "study design" "sample size" and "inclusion and exclusion criteria". Its aim is to promote the full understanding and use of the ARRIVE 2.0 guidelines by domestic researchers, enhance the standardization of experimental animal research and reporting, and promote the high-quality development of experimental animal technology and comparative medicine research in China.

**[Key words]** Animal experiment; ARRIVE 2.0 guidelines; ARRIVE essential 10; Study design; Sample size; Inclusion and exclusion criteria

透明准确的报告对于提高科学研究的可重复性至关重要；它使其他研究者能够仔细审查研究方法的严谨性，评估结果的可靠性，并能重复或借鉴其工作。然而，有证据表明，大多数出版物中缺少关键信息，涉及动物研究的报告还有极大的改进空间<sup>[1-4]</sup>。为此，英国国家3Rs中心（National Centre for the Replacement, Refinement and Reduction of Animals in Research, NC3Rs）在2010年发布了ARRIVE指南即《动物研究：体内实验报告》（*Animal Research: Reporting In Vivo Experiments*）。该指南罗列了文稿中应有信息的条目清

单，以确保出版物包含足够的信息，从而补充完善知识库<sup>[5]</sup>。该指南目前已被1 000多种期刊推荐使用，并获得了世界各地的研究资助者、大学和学术团体的广泛认可。

然而，目前对ARRIVE指南影响报告质量的评估结果不一<sup>[6-11]</sup>。有证据表明，从事体内研究的科学家尚没有充分意识到完整报告ARRIVE指南所涵盖信息的重要性，也没有意识到这些信息与其工作或研究领域的相关性<sup>[12]</sup>。作为一个新的国际工作组——该指南的作者们对第一版的ARRIVE指南进行了修订和更新，

以使其更易被采用；修订后的 2.0 版 ARRIVE 指南<sup>[13]</sup>及其完整解读版<sup>[14]</sup>（即本文的英文原稿）一起发布。作者们依据当前的最佳实践更新了诸多建议，并对信息进行了重构，将报告条目分为“关键 10 条”和“推荐 11 条”两个集合。ARRIVE 2.0 中的“关键 10 条”构成了研究报告的最低要求，“推荐 11 条”则为所述的研究提供了更多的背景内容。诚然，研究报告最好能同时包含这两部分条目的全部信息，但首先应关注关键问题，这有助于论文作者、期刊工作人员、编辑以及审稿人在实际工作中使用该指南，从而使指南得以切实地实施。一旦“关键 10 条”在所有稿件中得到一致报告，那么“推荐 11 条”就可以被逐步添加到期刊要求中，直到所有 21 项条目在所有论文中都被例行报告。NC3Rs 网站中描述了修订和将条目分组的完整方法<sup>[13]</sup>。

此次指南修订的一个重要举措是撰写这篇解释和阐明性文章，从而为 ARRIVE 2.0 指南中的 21 项条目（包括关键 10 条和推荐 11 条）提供背景和理论基础。为此，ARRIVE 2.0 指南的作者们为每项条目及其子条目提供了附加指南，进一步阐释了在描述动物研究文稿时报告这些信息的重要性及其支持性证据，详细说明了应该报告哪些内容及细节。该指南适用于生物科学研究中涉及活体动物的所有研究领域，包括哺乳动物以及模式生物（例如果蝇或秀丽隐杆线虫）。无论是单一动物研究的文稿，还是体内观察联合其他类型实验研究的范围更广的文稿，每项条目都同等重要。当然，要报告的具体细节类型可能因物种和实验设置而不同，因此需要在每个条目的附加指引中进一步确认。

ARRIVE 2.0 指南的作者们认识到，研究的目的会影响研究方案的设计。例如：验证性研究的目的是评估特定的假说，会采用严谨的实验方法降低实偏倚风险，并在研究启动之前制定详细的统计分析计划。相比之下，探索性研究经常同时研究多个问题，而不需要像验证性研究那样严格遵守严谨的标准；这种灵活性可用以开发或测试新的方法、产生某些可在以后正式验证的理论和假说。这两种研究类型都为科学进步做出了宝贵的贡献。准确透明地报告研究目的，以及在实验设计、实施和分析中采用的严谨性水平，有助于读者判断如何使用该项研究、研究结果是否具有开创性，以及在采纳借鉴该成果前是否需要进一步验证，或者这些研究是否足够可靠而能被应用于其他研究场景。

为了详细说明“关键 10 条”所述报告信息的重要性，此完整解读版还涵盖了实验设计的概念和最佳实践方案。这主要有两个目的：首先，有助于论文作者理解这些信息与读者评估研究报告结果的可靠性之间的相关性，因而鼓励作者全面详尽地报告相关信息；第二，它支持在动物研究设计和实施过程中践行最佳实践，即研究人员在计划进行体内实验伊始就查阅此完整解读版，将有助于他们充分利用 ARRIVE 2.0 指南，不仅将这些建议应用到研究设计中，而且在实验实施过程中收集相应的信息，以便为了遵循指南规范地报告研究结果。

为了确保这些建议对目标受众尽可能清晰和实用，作者们在撰写该指南时便与从事体内研究<sup>[13]</sup>的科研工作者一起，对 ARRIVE 2.0 指南及此完整解读版进行实用测试。每项条目都独立成章，有助于作者能够单独查阅特定条目内容；并且列出术语表以解释常见的统计学术语（表 1）。每个子条目也通过已发表文献中筛选出的优秀案例进行阐明释义，即给出最佳实践指南和建议，以提高研究人员从实验设计到论文发表的水平。解释说明部分和优秀案例均可从 ARRIVE 指南网站（<https://www.arriveguidelines.org>）获得。

## 1 ARRIVE 2.0 指南的“关键 10 条”

ARRIVE 2.0 关键 10 条是确保审稿人和读者能够评估研究结果可靠性的最低报告要求。这一部分中没有等级之分，10 个条目按逻辑顺序显示，包括：研究设计、样本量、纳入和排除标准、随机化、盲法、结果测量、统计方法、实验动物、实验步骤、结果<sup>[15]</sup>。

### 1.1 条目 1: 研究设计

每一项研究所需要提供的实验设计信息如下。

#### 1.1.1 子条目 1a: 比较的组别，含对照组。如果没有对照组，应阐明理由

详解：对照组或比较组的设置往往取决于实验目的。阴性对照用于确定组间差异是否由干预引起（例如：野生型和基因修饰动物，安慰剂组和积极治疗组，假手术组和手术干预组）。阳性对照是用于支持对阴性结果的解释或者确定预期的效果是否可以检测到。

有些情况下，可以不必单独设置非干预组，例如：当以比较不同给药方法（腹腔给药与经口灌服）为实验目的时，或在纵向研究中用作自身对照的动物。预实验有时也不需要设置对照组，例如：检验一个实验步骤或规程的可行性。

表1 ARRIVE 2.0指南所附统计学术语

Table 1 Statistical terminology attached to the ARRIVE 2.0 guidelines

| 术语名称<br>Terminology name                            | 含义<br>Content                                                                                                                                                               |
|-----------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 偏倚<br>Bias                                          | 对干预的真实效果的过高或过低估计。偏倚是由实验设计、实施或分析的不足引起,从而导致误差的引入                                                                                                                              |
| 统计描述和统计推断<br>Descriptive and inferential statistics | 统计描述用于总结数据,通常包括集中趋势(如平均值或中位数)的测量和离散程度(如标准差或范围)的测量。统计推断用于对从中抽取样本的总体进行概括。假设检验如方差分析(ANOVA)、Mann-Whitney 检验或 <i>t</i> 检验等属于统计推断的范畴                                              |
| 效应量<br>Effect size                                  | 组间差异或变量之间关系强度的定量测量                                                                                                                                                          |
| 实验单元<br>Experimental unit                           | 独立于所有其他单元而接受干预的生物实体,这样就可以将任何两个实验单元分配给不同的处理组。有时也被称为随机化单元。                                                                                                                    |
| 外部效度<br>External validity                           | 某一特定研究的结果能够应用或推广到其他研究、研究条件、动物品系/物种或人类的程度                                                                                                                                    |
| 假阴性<br>False negative                               | 当备择假设( $H_1$ )为真时,却得到无统计学意义的结果。在统计学中,它被称为第Ⅱ类错误                                                                                                                              |
| 假阳性<br>False positive                               | 当零假设( $H_0$ )为真时,却得到具有统计学意义的结果。在统计学中,它被称为第Ⅰ类错误                                                                                                                              |
| 自变量<br>Independent variable                         | 研究人员控制的变量(如处理、条件、时间)或样本的属性(如性别)或技术特征(如批次、笼、样本收集)等可能会影响结果测量的变量。自变量可以是科研中所关注的变量,也可以是干扰变量。自变量也被称为预测变量                                                                          |
| 内部效度<br>Internal validity                           | 某一特定研究的结果在多大程度上可以归因于实验干预的效果,而不是其他一些未知的因素(如该研究的设计、实施或分析的不足所引入的偏倚)                                                                                                            |
| 干扰变量<br>Nuisance variable                           | 干扰变量是指在研究中出现的一类变量,不是研究的主要关注点,但可能对研究结果测量产生影响并增加变异性,因此需要在实验设计或分析中加以考虑。此外,如果它们与关注的自变量有关联,就成为混杂因素,因为这会引入偏倚。在实验设计(以防止它们成为混杂因素)和分析(解释变异性,有时是降低偏倚)中都应对干扰变量加以考虑。例如,可将干扰变量作为区组因素或协变量 |
| 零假设和备择假设<br>Null and alternative hypotheses         | 零假设( $H_0$ )是指没有影响/效应,如各组之间的差异或变量之间的关联。备择假设( $H_1$ )是假设存在某种效应                                                                                                               |
| 结果测量/结局变量<br>Outcome measure                        | 在研究过程中记录的、用以评估处理或实验干预效果的任何变量。它也被称为因变量、响应变量                                                                                                                                  |
| 统计效力/检验效能<br>Power                                  | 对于预先定义的有生物学意义的效应量,如果效应真实存在(即零假设被正确拒绝),统计检验将检测出效应的概率                                                                                                                         |
| 样本量<br>Sample size                                  | 每组的实验单元数量,也被称为 <i>n</i>                                                                                                                                                     |

针对较为复杂的研究设计,推荐使用诸如时间轴、流程图等直观表述形式,前者比文字描述更容易理解。相比单纯的文字描述,图表更利于表述哪些处理和程序在哪个研究阶段应用于特定的动物或组别。图表还有助于表述复杂的设计特征,例如:多因素的交叉或嵌套(层次设计或多水平设计),区组设计(减少非必要的变异,详见条目4:随机化),或对某一个实验单元随时间进行反复检测的实验(重复测量设计);更多不同的实验设计类型,详见文献<sup>[16-18]</sup>。“实验设计助手”(The Experimental Design Assistant, EDA)是支持研究人员设计体内实验的辅助平台(网站: <https://eda.nc3rs.org.uk/>),可以生成图表来表示任何类型的实验

设计<sup>[19]</sup>。

每一个实验都应报告所有分组信息,包括实验组、对照组(阴性或阳性)等。选择性地排除一些实验组(例如,因为数据与论文的叙述不一致或有冲突)具有误导性,因此应避免<sup>[20]</sup>。确保能够清晰地区分实验组、比较组和对照组(阳性或阴性)。如果同一对照组用于多个实验或未设置对照组,需在报告中清晰地阐述。

报告示例1:“DAV1是一项单因素、两阶段交叉实验。实验所用药物为阿莫西林,受试动物为16头仔猪。该实验分为两个阶段:第一阶段,阿莫西林和安慰剂以混饲的方式饲喂仔猪;第二阶段,仅将含有阿

莫西林的饲料饲喂仔猪。其中，阿莫西林的口服剂量为 30 mg/kg（单一剂量）。在每一个实验阶段，分别采集仔猪采食后 0.5、1、1.5、2、4、6、8、10 和 12 h 的血浆以测定阿莫西林的浓度”<sup>[21]</sup>。

报告示例 2：“本研究方案的示意图由 EDA 设计完成。该示意图表明本研究是一个较为简单的对比研究，即两种药物处理两种不同的肿瘤细胞

系后，观察药物对肿瘤细胞转移、扩散的影响。采用区组随机化将斑马鱼胚胎随机等量分为 3 组。分组后，每枚胚胎注射两种肿瘤细胞中的一种，并对注射后的胚胎进行药物干预（包括空白对照组）。通过两因素方差分析，确定药物处理对每种肿瘤细胞系生长、存活和侵袭的效果”<sup>[22]</sup>。见图 1。

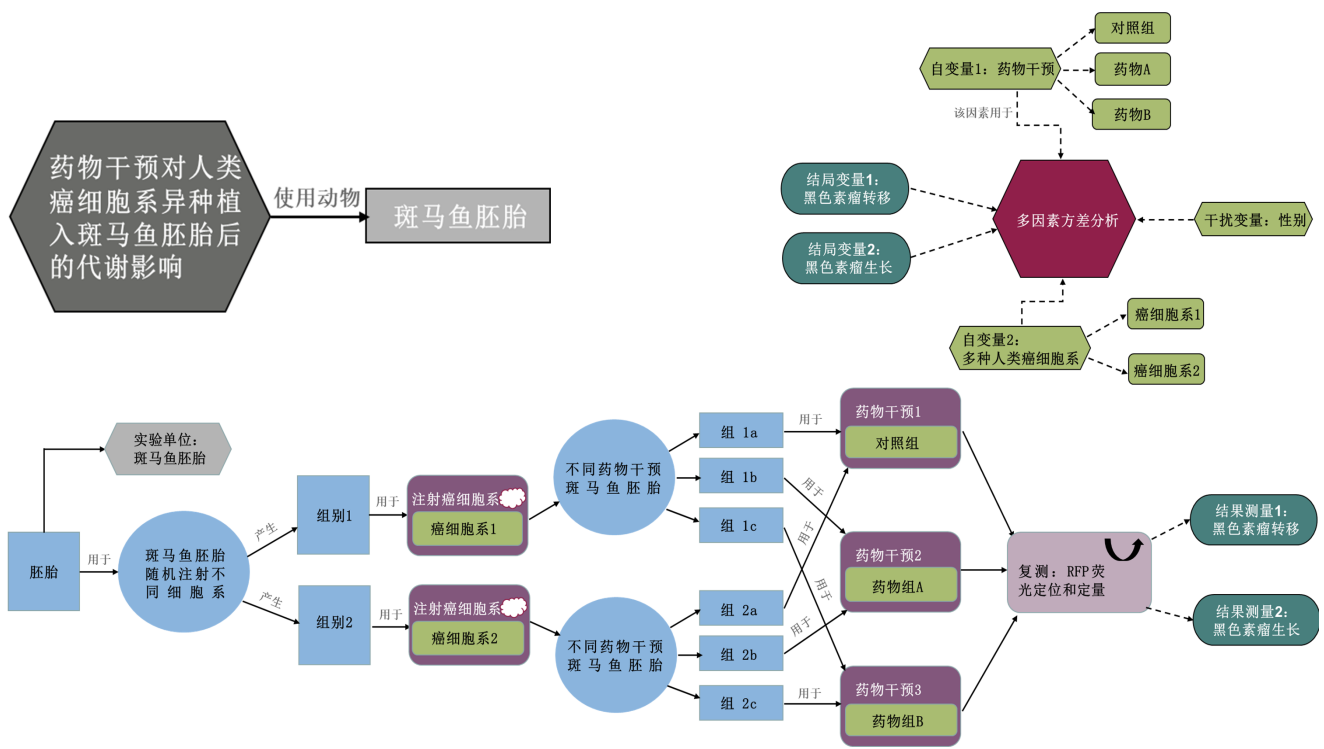


图1 文献[22]中“实验设计助手”设计的研究方案

Figure 1 The example of a study plan created by the Experimental Design Assistant (EDA) in study [22]

1.1.2 子条目 1b: 实验单元 (如: 以单只动物、一窝动物或一笼动物为单元)

详解：在一项实验设计中，其所涉及的生物性和技术性因素通常是分层次组织的，比如动物体内的细胞、细胞里的线粒体，或者房间里的笼盒、笼盒里的动物等。这样的层次会使确定样本量变得困难（是动物、细胞，还是线粒体的数量？）。样本量是每组的实验单元数。实验单元指的是独立于所有其他单元而接受干预的生物实体，这样就可以将任何两个实验单元分配给不同的处理组。有时也被称作随机单元。此外，实验单元不应相互影响所测量的结果。

通常而言，实验单元是被独立分配到处（例如，注射药物）组中的单个动物，也可以是一笼或一窝（例如，对一整笼的实验动物饲喂某种食物或者对母体

进行处理后研究其幼仔），也可以是动物身体的一部分（例如，将多种药物分别用于同一只动物身体的不同部位）。动物还可以做自身对照，接受不同的处理，其间由洗脱期分隔。此时，实验单元是指处于不同实验处理时期的单个动物。单个实验也可能会有多个实验单元，例如：先对妊娠雌性动物进行实验处理，然后对其离乳的幼仔分组饲喂不同的饲料<sup>[23]</sup>。关于实验单元的更多指南，详见文献<sup>[17, 24–25]</sup>。

将实验单元与子样本或重复测量相混淆会人为地导致样本量膨胀。例如：当实验单元是小鼠个体时，则来自同一只小鼠身上的 50 个细胞的测量值代表样本量为 1（即  $n=1$ ）。这 50 个测量值是子样本，提供测量误差的估计值，因此应取平均值或在巢式分析中使用。在这种情况下报告  $n=50$  就是一个典型的伪重复<sup>[26]</sup>，

这使得真实的变异性被低估,从而会导致假阳性和无效的分析结果<sup>[26-27]</sup>。然而,如果将取自小鼠身上的每个细胞随机分配到不同处理组内,且单独进行评估,那么每个细胞就可以作为实验单元。

每个实验都应明确指出其实验单元,以便对样本量和统计分析做出正确评估。

报告示例1:“本研究的样本为分别从饲喂含1倍胆碱和4倍胆碱日粮的妊娠15.5 d母鼠中采集的胎盘组织(每组每一胎鼠性别各有3只妊娠母鼠,共计12只妊娠小鼠)。为确保统计的独立性,每只妊娠小鼠只取一个胎盘(雄性或雌性胎鼠)用于实验。因此,胎盘即作为该实验的一个实验单元”<sup>[28]</sup>。

报告示例2:“我们使用了从高通量表型分析中收集的数据,该表型分析基于“流水线”或流程的形式进行表征,其中小鼠通过一系列以标准操作规程(SOP)为基础的标准化和验证性检测而被表征……这里的单只小鼠就是本研究中的实验单元”<sup>[29]</sup>。

报告示例3:“将鱼根据体质量(0.7~1.2 g和1.3~1.7 g)分为两组,并随机放养在24个装有60 L水的饲养容器内(每一个实验单元的密度为15条)”<sup>[30]</sup>。

报告示例4:“本研究中的 $n$ 是指动物数量,每组包括6只动物,每只动物最多获取3张皮质纹状体切片,每个切片采集5次数据”<sup>[31]</sup>。

## 1.2 条目2:样本量

1.2.1 子条目2a:详细说明分配给每个实验组的确切实验单元数量,以及每次实验的实验单元总数。同时说明整个实验使用的动物总数

详解:样本量大小与实验之初每个实验组包含的实验单元数量有关,通常用“ $n$ ”表示(详见条目1-研究设计,以获取关于实验单元的定义和报告准则的更多指南)。这些信息对于评估统计模型的有效性和实验结果的可靠性至关重要。

研究开始时每组的样本量可能与分析中的 $n$ 数量不同(详见条目3:纳入和排除标准)。此信息有助于读者了解实验的损耗,以及是否存在样本被排除的情况和这些情况出现在哪个组中。研究报告中对所用动物的总数进行说明,有助于识别实验中是否存在重复利用动物的情况。

报告每组准确的 $n$ 值,以及每个实验的总样本量(包括任何独立的重复)。如果实验单元不是动物,动物的总数同样需要阐述,以帮助读者更好地理解实验

设计方案。例如:在一项研究饲喂的实验中,以动物饲养“笼”为实验单元,并且每笼饲养一对动物,则所用动物总数是实验单元总数的两倍。

报告示例1:“实验中对每组动物实施的处理和移植方案。根据每组所接受的一期和二期处理对其进行命名。注:S为生理盐水;L为左旋多巴;SS组在一期和二期均给予了生理盐水;SL组在一期给予了生理盐水,在二期给予了左旋多巴;LS组在一期给予了左旋多巴,二期给予了生理盐水;LL组在两个阶段均给予了左旋多巴。实验开始时每组为12只动物,但有些动物出现了与实验无关的肿瘤,因此被从实验中移除”<sup>[32]</sup>。见图2。

| 处理组     | $n$ | 一期  | 移植类型 | 二期            |
|---------|-----|-----|------|---------------|
| 假手术组    | SS  | 6   | 生理盐水 | 假手术           |
|         | SL  | 8   | 生理盐水 | 假手术           |
|         | LS  | 8   | 左旋多巴 | 假手术           |
|         | LL  | 8   | 左旋多巴 | 假手术           |
| 同种异体移植组 | SS  | 10  | 生理盐水 | 14 d大鼠胚胎的腹侧中脑 |
|         | SL  | 11* | 生理盐水 | 14 d大鼠胚胎的腹侧中脑 |
|         | LS  | 12  | 左旋多巴 | 14 d大鼠胚胎的腹侧中脑 |
|         | LL  | 12  | 左旋多巴 | 14 d大鼠胚胎的腹侧中脑 |
| 异种异体移植组 | SS  | 10  | 生理盐水 | 12 d小鼠胚胎的腹侧中脑 |
|         | SL  | 10* | 生理盐水 | 12 d小鼠胚胎的腹侧中脑 |
|         | LS  | 11* | 左旋多巴 | 12 d小鼠胚胎的腹侧中脑 |
|         | LL  | 12  | 左旋多巴 | 12 d小鼠胚胎的腹侧中脑 |

注:\*表示这些组最初有12只动物,但一些动物患上了与实验无关的肿瘤,因此被从研究中移除。

Note: \*These groups initially consisted of 12 animals, but some animals developed tumors unrelated to the experiment and were therefore removed from the study.

图2 文献[32]中每组动物的处理和移植方案

Figure 2 The scheme of treatment and transplantation for each group of animals in study [32]

1.2.2 子条目2b:解释样本量是如何确定的。如已计算样本量,需提供先前计算的所有细节

详解:任何类型的实验报告都需要描述样本量大小的决定依据。对于确证性实验,由于采用统计推断来估计效应量大小,并确定推翻零假设的证据权重,因此需要证明样本量大小的合理性,以确保实验具有最佳样本量来检验研究问题<sup>[33-34]</sup>(详见条目13即推荐11条中的第3条:研究目标)。如果样本量过小(即检

验效力不足的研究), 则会产生不确定的结果, 而样本量过大 (即检验效力过强的研究) 则会因不必要的使用动物而引发伦理问题, 并可能产生虽具有统计学意义但与生物学无关的微不足道的结果<sup>[35]</sup>。检验效力低有3种影响: 首先, 在实验中, 真实的效应很有可能被遗漏; 其次, 当检测到效应时, 往往会是真实效应量的高估<sup>[24]</sup>; 最后, 当低检验效力与发表偏倚同时存在, 发表的文章中假阳性率就会增加<sup>[36]</sup>。因此, 低效力的实验会使研究的内部效度变差, 同时也在无定论的研究中有浪费动物的伦理风险<sup>[37]</sup>。

研究设计方案会影响实验的统计效力, 效力的计算需要与所实施的研究设计方案相匹配。针对各种实验设计和统计分析, 都已有协助进行事先样本量计算的统计程序, 既有免费资源 (在线小程序以及R语言中的功能), 也有付费的商业软件<sup>[38-40]</sup>。需要根据结局指标和自变量类型以及分组数量来选择合适的软件或算法。建议咨询统计学专家, 尤其是在设计复杂或不常见的实验设计类型时。

当实验用于测试干预措施对一个连续结果变量的平均值产生的影响时, 可以根据预先定义的与生物学相关的效应大小、依据先前数据预估的变异性、选定的显著水平、检验效能等因素之间的数学关系来计算样本量 (见下一段和Bate等<sup>[17]</sup>、Festing等<sup>[41]</sup>的实用性建议)。如果您事先进行了样本量计算, 请报告: (1) 统计分析方法 (例如: 采用双侧 $t$ 检验, 差异显著性 $P$ 值为0.05); (2) 所关注的效应量, 并解释选择该效应量大小的理由; (3) 变异性的估计指标 (例如标准差) 以及如何估算的; (4) 选择的检验效能。

效力计算时所用到的信息包括主要实验目的、预先确定的效应量、变异性估计、显著性阈值、效力、方向性。样本量的计算基于效应量大小、变异性、显著性水平、检验效能和样本量之间的数学关系。需要考虑以下问题:

(1) 主要的结局指标是什么? 应在实验的计划阶段就确定好主要结局指标是什么, 这是最重要的结果, 它将回答主要的实验问题。

(2) 具有生物学关联的效应量是什么? 效应量是指不同实验组之间主要结局指标的生物学相关性变化。可以参考类似的研究获得相关信息, 从而使研究人员去探索效应量达到何种大小时会具有科学意义, 并值得进一步开展工作。在临床前研究中, 还应考虑效应的临床相关性。

(3) 变异性如何估计? 变异性估计可以通过以下方面获得: 1) 根据与计划开展的实验条件完全相同的预实验所获得的数据来估计。例如: 以前在同一实验室所做的实验和类似条件下对具有相同特征的动物进行相同处理的实验。2) 从以前测试不同处理方法的实验中的对照组获得。3) 参照文献中报告的类似实验。

(4) 多大的假阳性风险是可以接受的? 显著性水平或阈值 ( $\alpha$ ) 是假阳性结果的发生概率。如果它被设定为0.05, 那么对于单次统计检验来说, 获得假阳性的概率在1/20。然而, 在多重检验中, 需要对显著性阈值或 $P$ 值进行调整 (例如: 使用Bonferroni校正)。

(5) 多大的假阴性风险是可以接受的? 对于一个预先定义的具有生物学意义的效应值, 检验效能 (即 $1-\beta$ ) 是统计检验可以检测到一个真实存在的效应 (即真阳性结果) 的概率。通常80%~95%的目标效力被认为是可接受的, 而这意味着还有5%~20%的假阴性风险。

(6) 使用单侧检验还是双侧检验?

检验的方向性取决于特定分析中测试数据的分布。对于基于 $t$ 或 $z$ 分布的检验 (如 $t$ 检验), 是否使用单侧或双侧检验来分析数据, 与备择假设是否具有方向性有关。一个具有方向性 (单侧) 备择假设的实验, 可以用单侧检验进行分析, 从而最大限度地提高检测该效应的灵敏度。在统计学领域, 对于何时使用单侧检验存在一定的争议<sup>[42]</sup>。单侧检验的使用需要证明为什么只有在某个设定方向上的处理效应才有意义, 以及为什么将另一个方向上的大的效应当作非显著性差异对待<sup>[43]</sup>。在使用单侧检验后, 研究者就无法再检验另一个方向上是否有遗漏效应的可能。仅仅为了获得统计学上的显著性差异而选择单侧检验是不合适的。具有无方向性备择假设的双侧检验更为常见, 它允许研究人员在不考虑方向性的情况下, 检验实验处理的效果。需要注意的是, 诸如方差分析 (ANOVA) 和卡方检验是基于只有一尾的不对称分布 ( $F$ 分布和卡方分布)。因此, 这些检验没有方向性选项, 即只能是单侧检验。

有几种类型的研究不适合事先进行样本量的计算。例如, 生产抗体或组织所需的动物数量取决于所需生产的数量以及单个动物的产出能力。对于以成功获得样本或状况 (例如, 获得转基因动物) 为结果的实验, 动物数量是由实验过程的成功率决定的。

在早期的可行性研究或预实验中, 动物的需求量

取决于研究目的。当初步研究的主要目标是流程或操作性质时（例如，为了改进程序和设备），所需的动物数量一般较少。在这种情况下，就不宜使用效力计算，样本量的大小可以根据操作能力和限制条件进行估计<sup>[44]</sup>。仅靠预实验不太可能为未来实验的效力计算提供足够的变异性数据。系统综述和先前的研究是变异性信息的更合适来源<sup>[45]</sup>。

如果没有使用效力计算来确定样本量，需要在报告中对此进行说明，并提供决定每组样本量的理由。无论是否使用了效力计算，在解释如何确定样本量时，都需要考虑到任何可能出现的动物或数据损失，例如，根据事先确定好的排除标准或预期损耗（详见条目3：纳入和排除标准）。

报告示例1：“样本量大小是根据参考治疗组即丁丙诺啡给药后的术后疼痛数字评分量表（NRS）得分（NRS AUC 平均值=2.70；非劣效性界值=0.54；标准偏差=0.66）……以及格拉斯哥综合疼痛量表（GCPs）……使用在线软件（实验设计助手：<https://eda.nc3rs.org.uk/eda/login/auth>）进行计算的。检验效能被设定为80%。每组至少需要20只犬”<sup>[46]</sup>。

报告示例2：“本研究首次对生物玻璃原型进行了体内实验评估，故采用了一个较小的样本量。因此，研究的主要目的是收集基本实验数据，为今后更复杂的实验设计提供数据支持”<sup>[47]</sup>。

### 1.3 条目3：纳入和排除标准

1.3.1 子条目3a：描述实验期间用于纳入和排除动物（或实验单元），以及分析过程中纳入和排除数据点的所有标准。详细说明这些标准是否为预先设定。如未设定相关标准，应给予明确声明

详解：纳入和排除标准定义了实验开始后实验动物及数据被采用与否的筛选条件。因此，为了确保科学上的严谨性，应在开展实验和收集数据之前制定这些标准<sup>[8, 33, 48-49]</sup>。纳入标准不应与动物特征相混淆（详见条目8：实验动物），但可能与之相关（例如：对某特定操作程序而言，体质量必须在特定范围内），或与其他研究指标有关（例如：任务表现必须超过既定的阈值）。在为不同目的选择数据进行重新分析的研究中，纳入和排除标准应描述如何选择数据。

技术或动物福利问题，是制定排除标准时所需要考虑的重要因素，例如：手术期间预期的并发症或其他可能会影响实验程序的情况（如运动损伤的发生有可能影响动物行为学检测结果）。实验样本或数据的排

除标准包括不符合相关质量控制标准，比如样品量不足、受到污染程度不可接受、组织质量欠佳等。同样，研究人员在分析过程中如何定义和处理异常数据也应在正式实验开始之前确定（详见子条目3b：可靠的数据清理指南）。

排除标准还可以反映出一项研究的伦理原则和其人道终点相一致（详见条目16即推荐11条的第6条：动物照护和监测）。例如：在癌症相关研究中，如果动物皮下肿瘤的大小超过规定体积，那么该动物可能在预设的时间点之前从研究中剔除并实施安乐死<sup>[50]</sup>。如果动物数量的损失可以预计，在决定实验入组的动物数量时应该考虑这些因素（详见条目2：样本量）。尽管排除标准和人道终点通常包含在实验伦理审查申请中，但是同时在文稿中报告用于排除实验动物或相关数据的标准将有助于读者解读实验数据，并为其他想采用该模型的研究人员提供关键信息。

最佳的做法是在预注册的研究方案中包含所有事先设定的纳入和排除标准，以及异常值的判定标准（详见条目19及推荐11条中的第9条：研究方案注册）。至少应在实验记录本中记录纳入和排除标准并在稿件中报告，明确指出在收集任何数据之前就定义了相关标准。

报告示例1：“如果这些动物成功地接受了大脑动脉栓塞（MCAo），则被纳入研究。MCAo的定义是通过激光多普勒血流仪观察到动物大脑血流量下降60%或更多。如果线栓导致血管壁穿孔（根据动物安乐死时蛛网膜下腔有出血来判断），或线栓的硅胶头在撤线过程中脱落，或动物过早死亡导致无法采集行为和组织学数据，则从研究中排除”<sup>[51]</sup>。

1.3.2 子条目3b：对于每个实验组，报告分析中排除的任何动物、实验单元或数据点，并说明原因。如果没有排除的情况，也请说明

详解：未予说明的动物、实验单元以及数据点可能导致原始数据无法支持研究结论<sup>[52]</sup>。对排除和损耗进行报告为其他研究人员评估实验结果、重复实验或在其他物种上测试该干预措施提供了有价值的信息，也可以为人体试验提供重要的安全性信息（例如，与不良反应相关的排除）。

实验损耗有很多合理的原因，其中有些是可预期并可预先进行控制的（详见子条目3a：如何定义排除和纳入标准），但是有些数据的缺失是不可预期的，例如：因为动物接受了错误的处理、非预期的药物毒性、

与实验无关的感染或疾病、采样错误（例如：因检测失误导致虚假结果，及设备校准不当），或其他人为错误（例如，忘记打开记录设备）。

大多数的统计分析方法对异常值和缺失值极为敏感。在某些情况下，剔除异常数据可能是科学合理的，比如数据输入中的明显错误或读数超出合理范围的测量。然而，不恰当的数据清理可能使研究结果产生偏倚<sup>[53]</sup>，提供数据剔除的理由可以有效地区分可靠的数据清理与数据操纵。数据缺失普遍存在于各个研究领域，如果缺失值不是随机的，则可能会影响研究的敏感性，还可能导致估计偏差、效力失真和信息丢失<sup>[54]</sup>。分析方案应包括探索数据丢失原因的方法。针对缺失数据调整统计分析方法同样重要<sup>[55-56]</sup>。

随着共享代码以实现分析重现等策略的增加，数据开放共享得以持续发展（详见条目20即推荐11条中的第10条：数据获取）。虽然这些做法相对公开透明，但仍然需要进一步公布数据清理的合理性，以及数据清理方法是否在收集数据之前定义。

应报告所有动物排除和数据点缺失，以及排除它们的理由。例如：以上信息可以用表格或流程图的形式加以概括，以描述每个组别的缺失情况。同时还须说明在排除数据或动物时研究者是否对分组情况不知情（详见条目5-盲法，以及Holman等<sup>[57]</sup>的研究），并明确说明剔除异常值所使用的软件和统计模型（例如，GraphPad Prism的异常值检验）。

报告示例1：“对所有数据而言，动物饲养围栏（简称畜栏）是实验单元。该实验的目的是通过对不同畜栏中饲养的肉公牛饲喂含有不同比例盐酸莱克多巴胺（RAC）和锌氨基酸螯合物（ZnAA）添加剂的饲料，观察不同组别间肉公牛的热胴体重。在实验当中，有一个畜栏内的肉公牛（饲喂90 mg/kg ZnAA）出现了与实验本身无关的疾病而表现不佳，故将该畜栏所有数据从实验前（Pre-RAC）和实验期间的数据中全部剔除……”。“通过使用Cook's *D* 统计值确定异常值，若Cook's *D* 值 > 0.5，则将其删除。有一头肉公牛饲喂48 d的肝脏活检微量无机物（TM）结果异常，故相关数据被删除”<sup>[58]</sup>。

报告示例2：“72只自发性高血压大鼠（SHR）被随机分配到实验当中，其中13只不符合我们的纳入和排除标准：7只的脑血流量下降未达到60%；1只术后死亡（尸检无法确定死因）；1只死于线栓介入过程中大出血；4只在线栓撤出时，涂在尼龙线远端的硅铜弹

性体脱落，使得再灌注的持久性不确定。因此，共有59只动物纳入本研究中梗死体积的分析。因为失误，有3只实验动物在神经行为学评分的最终评估之前被安乐死，其中1只来自常温/水组，两只来自低温/哌替啶组。这些失误发生时，人员对处理组分配不知情。所以，最后只有56只SHR大鼠被纳入神经行为学评分分析中”<sup>[51]</sup>。

报告示例3：“流程图显示了实验方案以及研究中所使用、死亡和纳入的动物数量……在心血管磁共振成像（CMR）和超声心电图的基础数据确立后，如前所述，通过左前降支（LAD）冠状动脉结扎术（*n*=48）诱发了心肌梗死（MI）。作为手术操作的对照，假手术组小鼠接受了开胸术和心包切开术，但没有进行冠状动脉结扎（*n*=12）”<sup>[59]</sup>。见图3。

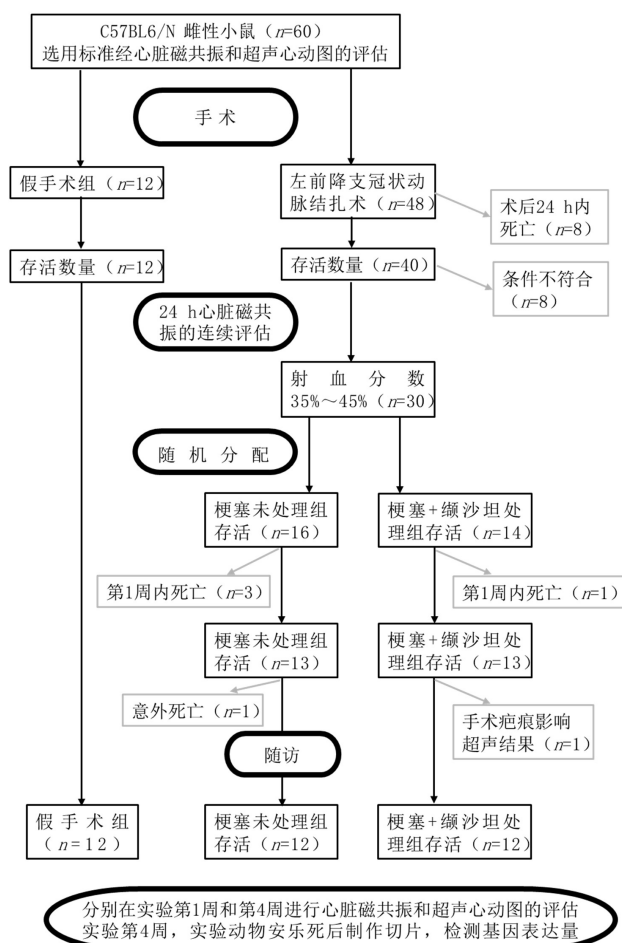


图3 文献 [59] 中实验方案及其所使用、死亡和纳入的动物数量

Figure 3 Flow chart showing the experimental protocol with the number of animals used, died and included in the study [59]

### 1.3.3 子条目 3c: 对于每次分析, 报告每个实验组中被纳入分析的动物、实验单元或数据点的确切数量( $n$ )

详解: 每组中被分析的实验单元的确切数量(即 $n$ 值)是读者解读分析报告时所需的重要信息, 因此应明确报告。实验中所有使用过的动物及其数据都应在报告中得到应有的体现。有时出于充分的理由, 可能需要将动物排除在研究之外(比如疾病或死亡), 或者把数据从分析中剔除(比如生物学上不合理的数值等)。有关动物缺失的报告, 有助于读者更好地理解实验设计过程并重复该方法, 有助于全程追踪研究中的动物数量, 尤其是当分析中的样本量与原始分组数量不匹配时。

对于每个研究结局, 应在文本或图表中清楚地标明数字并提供绝对数字(例如 10/20, 而不是 50%)。对于需要在不同时间点对实验动物进行检测的研究, 报告中需明确阐述哪些动物在何时接受检测<sup>[33]</sup>。

报告示例 1: “F 组在 2010 年包含 29 只成年雄性和 58 只成年雌性恒河猴 ( $n=87$ ), 在 2011 年则包括 32 只成年雄性和 66 只成年雌性恒河猴 ( $n=98$ )。雌性数量的增加是由于从青少年逐渐发育成熟至成年。雌性属于 3 个母系族群, 而雄性的等级没有大的变化。6 个中低等级的个体因死亡而未纳入分析, 2011 年初从该组移出的 5 个中等级别的雄性也未纳入分析”<sup>[59]</sup>。

报告示例 2: “通过视频记录动物与饲养员之间互动时所表现出的动作及时长, 比如闻戴手套的手或管道、伸出爪子接触、攀爬或进入管道, 并统计互动时间占总时间的比例。然后对每笼中的 2 只小鼠取平均值, 因为它们同时进行测试, 彼此的行为并非独立。该实验发现: 相比那些用抓尾方式抓取的小鼠 ( $6.4\%\pm 2.0\%$ , 测试时间 60 s,  $n=8$  笼), 使用原来笼具里的管道辅助抓取 ( $39.8\%\pm 5.2\%$ , 60 s,  $n=8$  笼) 的小鼠与操作人员互动更多。而那些采用手捧方式抓取的实验组小鼠与人员的互动水平介于上述两种方法之间 ( $27.6\%\pm 7.1\%$ ,  $n=8$  笼)”<sup>[60]</sup>。

(待续)

#### [作者贡献 Author Contribution]

王剑负责 ARRIVE 2.0 指南完整版第一部分即摘要、前言和“关键 10 条”前 3 条内容的编译初稿和修改;

卢今、马政文参与 ARRIVE 2.0 指南完整版第一部分内容的编译初稿写作;

陈国元、卢晓、白玉、刘晓宇和卢选成负责对 ARRIVE 2.0 指南完整版第一部分编译内容提出关键性修改建议;

高静负责 ARRIVE 2.0 指南完整版中统计学概念及阐述的翻译审定;

李焱负责 ARRIVE 2.0 指南完整版中文编译工作的策划、组织, 以及第一部分内容的初稿审定和监督指导;

庞万勇负责 ARRIVE 2.0 指南完整版中文编译工作的指导, 以及全部内容的审定。

#### [利益声明 Declaration of Interest]

本文所有作者均声明不存在利益冲突。

#### [参考文献 References]

- [1] KILKENNY C, PARSONS N, KADYSZEWSKI E, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals[J]. PLoS One, 2009, 4(11): e7824. DOI: 10.1371/journal.pone.0007824.
- [2] FLÓREZ-VARGAS O, BRASS A, KARYSTIANIS G, et al. Bias in the reporting of sex and age in biomedical research on mouse models[J]. eLife, 2016, 5: e13615. DOI: 10.7554/eLife.13615.
- [3] WEISSGERBER T L, GARCIA-VALENCIA O, GAROVIC V D, et al. Why we need to report more than 'Data were Analyzed by t-tests or ANOVA'[J]. eLife, 2018, 7: e36163. DOI: 10.7554/eLife.36163.
- [4] MACLEOD M R, LAWSON MCLEAN A, KYRIAKOPOULOU A, et al. Correction: risk of bias in reports of *in vivo* research: a focus for improvement[J]. PLoS Biol, 2015, 13(11): e1002301. DOI: 10.1371/journal.pbio.1002301.
- [5] KILKENNY C, BROWNE W J, CUTHILL I C, et al. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research[J]. Osteoarthritis Cartilage, 2012, 20(4):256-260. DOI: 10.1016/j.joca.2021.02.010.
- [6] BAKER D, LIDSTER K, SOTTOMAYOR A, et al. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies[J]. PLoS Biol, 2014, 12(1): e1001756. DOI: 10.1371/journal.pbio.1001756.
- [7] GULIN J E N, ROCCO D M, GARCÍA-BOURNISSEN F. Quality of reporting and adherence to ARRIVE guidelines in animal studies for chagas disease preclinical drug research: a systematic review[J]. PLoS Negl Trop Dis, 2015, 9(11): e0004194. DOI: 10.1371/journal.pntd.0004194.
- [8] AVEY M T, MOHER D, SULLIVAN K J, et al. The devil is in the details: incomplete reporting in preclinical animal research [J]. PLoS One, 2016, 11(11): e0166733. DOI: 10.1371/journal.pone.0166733.
- [9] REICHLIN T S, VOGT L, WÜRBEL H. The researchers' view of scientific rigor-survey on the conduct and reporting of *in vivo* research[J]. PLoS One, 2016, 11(12): e0165999. DOI: 10.1371/journal.pone.0165999.
- [10] LEUNG V, ROUSSEAU-BLASS F, BEAUCHAMP G, et al. ARRIVE has not ARRIVED: support for the ARRIVE (Animal Research: reporting of *in vivo* Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia[J]. PLoS One, 2018, 13(5): e0197882. DOI: 10.1371/journal.pone.0197882.
- [11] HAIR K, MACLEOD M R, SENA E S, et al. A randomised

- controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus)[J]. *Res Integr Peer Rev*, 2019, 4:12. DOI: 10.1186/s41073-019-0069-3.
- [12] MCGRATH J C, LILLEY E. Implementing guidelines on reporting research using animals (ARRIVE etc.): new requirements for publication in BJP[J]. *Br J Pharmacol*, 2015, 172(13):3189-3193. DOI: 10.1111/bph.12955.
- [13] PERCIE DU SERT N, HURST V, AHLUWALIA A, et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research[J]. *BMC Vet Res*, 2020, 16(1):242. DOI: 10.1186/s12917-020-02451-y.
- [14] DU SERT N P, AHLUWALIA A, ALAM S, et al. Reporting animal research: explanation and elaboration for the ARRIVE guidelines 2.0[J]. *PLoS Biol*, 2020, 18(7): e3000411. DOI: 10.1371/journal.pbio.3000411.
- [15] 张俊彦, 刘晓宇, 李焱, 等. 动物实验研究报告的国际指南 ARRIVE 2.0 介绍及期刊实施计划[J]. *实验动物与比较医学*, 2023, 43(1): 86-94. DOI: 10.12300/j.issn.1674-5817.2023.014.
- ZHANG J Y, LIU X Y, LI Y, et al. Introduction to the international guide for animal research reporting ARRIVE 2.0, and its implementation plan in the journal[J]. *Lab Anim Comp Med*, 2023, 43(1): 86-94. DOI: 10.12300/j.issn.1674-5817.2023.014.
- [16] FESTING M F W, ALTMAN D G. Guidelines for the design and statistical analysis of experiments using laboratory animals [J]. *ILAR J*, 2002, 43(4):244-258. DOI: 10.1093/ilar.43.4.244.
- [17] BATE S T, CLARK R A. The design and statistical analysis of animal experiments[M]. Cambridge: Cambridge University Press, 2014.
- [18] RUXTON G D, COLEGRAVE N. Experimental design for the life sciences[M]. Oxford: Oxford University Press, 2017.
- [19] PERCIE DU SERT N, BAMSEY I, BATE S T, et al. The experimental design assistant[J]. *PLoS Biol*, 2017, 15(9): e2003779. DOI: 10.1371/journal.pbio.2003779.
- [20] The BMJ. Scientific misconduct [EB/OL]. [2020-1-10]. <https://www.bmj.com/about-bmj/resources/authors/forms-policies-and-checklists/scientific-misconduct>.
- [21] NGUYEN T T, BAZZOLI C, MENTRÉ F. Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models[J]. *Stat Med*, 2012, 31(11-12): 1043-1058. DOI: 10.1002/sim.4390.
- [22] HILL D, CHEN L P, SNAAR-JAGALSKA E, et al. Embryonic zebrafish xenograft assay of human cancer metastasis[J]. *F1000Research*, 2018, 7:1682. DOI: 10.12688/f1000research.16659.2.
- [23] BURDGE G C, LILLYCROP K A, JACKSON A A, et al. The nature of the growth pattern and of the metabolic response to fasting in the rat are dependent upon the dietary protein and folic acid intakes of their pregnant dams and post-weaning fat consumption[J]. *Br J Nutr*, 2008, 99(3): 540-549. DOI: 10.1017/S0007114507815819.
- [24] LAZIC S E, CLARKE-WILLIAMS C J, MUNAFÒ M R. What exactly is 'N' in cell culture and animal experiments?[J]. *PLoS Biol*, 2018, 16(4): e2005282. DOI: 10.1371/journal.pbio.2005282.
- [25] NC3Rs. Experimental unit 2015 [EB/OL]. [2019-3-21]. <https://eda.nc3rs.org.uk/experimental-designunit>.
- [26] LAZIC S E. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis?[J]. *BMC Neurosci*, 2010, 11:5. DOI: 10.1186/1471-2202-11-5.
- [27] HURLBERT S H. Pseudoreplication and the design of ecological field experiments[J]. *Ecol Monogr*, 1984, 54(2):187-211. DOI: 10.2307/1942661.
- [28] KWAN S T C, KING J H, GRENIER J K, et al. Maternal choline supplementation during normal murine pregnancy alters the placental epigenome: results of an exploratory study[J]. *Nutrients*, 2018, 10(4):417. DOI: 10.3390/nu10040417.
- [29] KARP N A, MASON J, BEAUDET A L, et al. Prevalence of sexual dimorphism in mammalian phenotypic traits[J]. *Nat Commun*, 2017, 8:15475. DOI: 10.1038/ncomms15475.
- [30] RIBEIRO F, VASQUEZ L A, FERNANDES J B K, et al. Feeding level and frequency for freshwater angelfish[J]. *R Bras Zootec*, 2012, 41(6): 1550-1554. DOI: 10.1590/S1516-35982012-000600033.
- [31] GRASSELLI G, ROSSI S, MUSELLA A, et al. Abnormal NMDA receptor function exacerbates experimental autoimmune encephalomyelitis[J]. *Br J Pharmacol*, 2013, 168(2): 502-517. DOI: 10.1111/j.1476-5381.2012.02178.x.
- [32] LUDIVINE S, BREGER, . Influence of chronic L-DOPA treatment on immune response following allogeneic and xenogeneic graft in a rat model of Parkinson's disease[J]. *Brain Behav Immun*, 2017, 61:155-164. DOI: 10.1016/j.bbi. 2016. 11.014.
- [33] VAHIDY F, SCHÄBITZ W R, FISHER M, et al. Reporting standards for preclinical studies of stroke therapy[J]. *Stroke*, 2016, 47(10):2435-2438. DOI: 10.1161/STROKEAHA.116.013643.
- [34] MUHLHAUSLER B S, BLOOMFIELD F H, GILLMAN M W. Whole animal experiments should be more like human randomized controlled trials[J]. *PLoS Biol*, 2013, 11(2): e1001481. DOI: 10.1371/journal.pbio.1001481.
- [35] JENNIONS M D, MØLLER A P. A survey of the statistical power of research in behavioral ecology and animal behavior [J]. *Behav Ecol*, 2003, 14(3): 438-445. DOI: 10.1093/beheco/ 14.3.438.
- [36] BUTTON K S, IOANNIDIS J P A, MOKRYSZ C, et al. Power failure: why small sample size undermines the reliability of neuroscience[J]. *Nat Rev Neurosci*, 2013, 14(5): 365-376. DOI: 10.1038/nrn3475.
- [37] WÜRBEL H. More than 3Rs: the importance of scientific validity for harm-benefit analysis of animal research[J]. *Lab Anim (NY)*, 2017, 46(4):164-166. DOI: 10.1038/labani.1220.
- [38] TEAM R. R: a language and environment for statistical computing[M]. Vienna, Austria: R Foundation for Statistical Computing, 2017.
- [39] PENG C Y, LONG H Y, ABACI S. Power analysis software for educational researchers[J]. *J Exp Educ*, 2012, 80:113-136. DOI: 10.1080/00220973.2011.647115.
- [40] CHARAN J, KANTHARIA N D. How to calculate sample size in animal studies?[J]. *J Pharmacol Pharmacother*, 2013, 4(4):303-306. DOI: 10.4103/0976-500X.119726.

- [41] FESTING M F. On determining sample size in experiments involving laboratory animals[J]. Lab Anim, 2018, 52(4):341-350. DOI: 10.1177/0023677217738268.
- [42] FREEDMAN L S. An analysis of the controversy over classical one-sided tests[J]. Clin Trials, 2008, 5(6):635-640. DOI: 10.1177/1740774508098590.
- [43] RUXTON G D, NEUHÄUSER M. When should we use one-tailed hypothesis testing?[J]. Methods Ecol Evol, 2010, 1(2):114-117. DOI: 10.1111/j.2041-210X.2010.00014x.
- [44] REYNOLDS P S. When power calculations won't do: Fermi approximation of animal numbers[J]. Lab Anim (NY), 2019, 48(9):249-253. DOI: 10.1038/s41684-019-0370-2.
- [45] Bate ST. How to decide your sample size when the power calculation is not straightforward [EB/OL]. [2018-8-2]. <https://www.nc3rs.org.uk/news/howdecide-your-sample-size-when-power-calculation-not-straightforward>.
- [46] BUSTAMANTE R, DAZA M A, CANFRÁN S, et al. Comparison of the postoperative analgesic effects of cimicoxib, buprenorphine and their combination in healthy dogs undergoing ovariohysterectomy[J]. Vet Anaesth Analg, 2018, 45(4):545-556. DOI: 10.1016/j.vaa.2018.01.003.
- [47] SPIN J, OLIVEIRA G, SPIN-NETO R, et al. Histomorphometric evaluation of the association between bioglass and lyophilized bovine bone in the treatment of critical bone defects created on rat calvaria: a pilot study[J]. Rev Odontol UNESP, 2015, 44(1):37-43. DOI: 10.1590/1807-2577.1020.
- [48] RICE A S C, MORLAND R, HUANG W, et al. Transparency in the reporting of *in vivo* pre-clinical pain research: The relevance and implications of the ARRIVE (Animal Research: Reporting In Vivo Experiments) guidelines[J]. Scand J Pain, 2013, 4(2):58-62. DOI: 10.1016/j.sjpain.2013.02.002.
- [49] SALKIND N J. Encyclopedia of research design[M]. Thousand Oaks, Calif: SAGE Publications, 2010.
- [50] WORKMAN P, ABOAGYE E O, BALKWILL F, et al. Guidelines for the welfare and use of animals in cancer research[J]. Br J Cancer, 2010, 102(11):1555-1577. DOI: 10.1038/sj.bjc.6605642.
- [51] SENA E S, JEFFREYS A L, COX S F, et al. The benefit of hypothermia in experimental ischemic stroke is not affected by pethidine[J]. Int J Stroke, 2013, 8(3):180-185. DOI: 10.1111/j.1747-4949.2012.00834.x.
- [52] KAFKAFI N, AGASSI J, CHESLER E J, et al. Reproducibility and replicability of rodent phenotyping in preclinical studies [J]. Neurosci Biobehav Rev, 2018, 87:218-232. DOI: 10.1016/j.neubiorev.2018.01.003.
- [53] SCOTT S, KRANZ J E, COLE J, et al. Design, power, and interpretation of studies in the standard murine model of ALS [J]. Amyotroph Lateral Scler, 2008, 9(1): 4-15. DOI: 10.1080/17482960701856300.
- [54] KANG H. The prevention and handling of the missing data[J]. Korean J Anesthesiol, 2013, 64(5): 402-406. DOI: 10.4097/kjae.2013.64.5.402.
- [55] DAVID ALLISON P. Missing data[M]. Thousand Oaks, Calif: Sage Publications, 2002.
- [56] JAKOBSEN J C, GLUUD C, WETTERSLEV J, et al. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts[J]. BMC Med Res Methodol, 2017, 17(1):162. DOI: 10.1186/s12874-017-0442-1.
- [57] HOLMAN C, PIPER S K, GRITTNER U, et al. Where have all the rodents gone? the effects of attrition in experimental research on cancer and stroke[J]. PLoS Biol, 2016, 14(1): e1002331. DOI: 10.1371/journal.pbio.1002331.
- [58] GENTHER-SCHROEDER O N, BRANINE M E, HANSEN S L. Effects of increasing supplemental dietary Zn concentration on growth performance and carcass characteristics in finishing steers fed ractopamine hydrochloride[J]. J Anim Sci, 2018, 96(5):1903-1913. DOI: 10.1093/jas/sky094.
- [59] CASTIGLIONI L, COLAZZO F, FONTANA L, et al. Evaluation of left ventricle function by regional fractional area change (RFAC) in a mouse model of myocardial infarction secondary to valsartan treatment[J]. PLoS One, 2015, 10(8): e0135778. DOI:10.1371/journal.pone.0135778.
- [60] BRENT L J N, HEILBRONNER S R, HORVATH J E, et al. Genetic origins of social networks in rhesus macaques[J]. Sci Rep, 2013, 3:1042. DOI: 10.1038/srep01042.
- [61] GOUVEIA K, HURST J L. Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling[J]. Sci Rep, 2017, 7: 44999. DOI: 10.1038/srep44999.

(收稿日期:2023-03-30 修回日期:2023-04-20)

(本文编辑:张俊彦,富群华)

#### [引用本文]

王剑, 卢今, 马政文, 等.《动物研究:体内实验报告》即 ARRIVE 2.0 指南的解释和阐述(一)[J]. 实验动物与比较医学, 2023, 43(2): 213-224. DOI: 10.12300/j.issn.1674-5817.2023.043.

WANG J, LU J, MA Z W, et al. Explanation and elaboration for the ARRIVE Guidelines 2.0—reporting animal research and *in vivo* experiments ( I ) [J]. Lab Anim Comp Med, 2023, 43(2): 213-224. DOI: 10.12300/j.issn.1674-5817.2023.043.