

医学院实验动物科学部, 上海 200025; 6. 中国疾病预防控制中心职业卫生与中毒控制所, 北京 100050; 7. 中国疾病预防控制中心, 北京 102206; 8. 上海交通大学医学院临床研究中心, 上海 200025; 9. 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评估该研究结果的可靠性, 进而重复该实验或在该成果的基础上进一步探索。ARRIVE 2.0 指南是英国国家 3Rs 中心 (NC3Rs) 于 2019 年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报告的细节, 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 (第二部分包括“关键 10 条”里的第 4~7 条: “随机化”、“盲法”、“结果测量”和“统计方法”等部分), 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; 随机化; 盲法; 结果测量;

统计方法

[中图分类号] R-332; Q95-33 **[文献标志码]** A **[文章编号]** 1674-5817(2023)03-0323-09



Explanation and Elaboration of the ARRIVE Guidelines 2.0—Reporting Animal Research and *In Vivo* Experiments (II)

CHEN Guoyuan^{1*}, LU Xiao^{2*}, BAI Yu^{3*}, YU Lingzhi⁴, QIAO Ying³, WANG Jian⁵, LU Jin⁵, LIU Xiaoyu⁶, LU Xuancheng⁷, GAO Jing⁸, LI Yao⁵, PANG Wanyong⁹

(1. Animal Core Facility, Center for Excellence in Molecular Cell Science, Chinese Academy of Science, Shanghai 200031, China; 2. Dizal Pharmaceuticals Co., Ltd., Shanghai 201203, China; 3. Novo Nordisk Research Centre China, Beijing 102206, China; 4. Shanghai Laboratory Animal Research Center, Shanghai 201203, China; 5. Department of Laboratory Animal Science, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; 6. National Institute for Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, Beijing 100050, China; 7. Laboratory Animal Center, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 8. Clinical Research Institute, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; 9. 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 remains a major challenge. Transparent and accurate reporting of progress can help readers evaluate the reliability of research results and further explore an experiment by repeating 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 the 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 the second part of the Chinese translation of the complete interpretation of the ARRIVE 2.0 guidelines published in *PLoS Biology* in 2020 (original text can be found at <https://arriveguidelines.org>) and based on the best practices for following the ARRIVE 2.0 guidelines in international journals. This part includes Items 4-7 of "ARRIVE Essential 10" in the ARRIVE 2.0 guidelines:

"Randomization", "Blinding", "Outcome Measurement", and "Statistical Methods". Our Chinese translated version aims to promote the full understanding and use of the ARRIVE 2.0 guidelines by domestic researchers, enhancing the standardization of experimental animal research and reporting, and promoting 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; Randomization; Blinding; Outcome measurement; Statistical methods

PLoS Biology 期刊于2020年发表了ARRIVE 2.0指南完整解读版（原文请见<https://arriveguidelines.org>）。该解读版是一份适用于任何与活体动物研究报告相关的，用以指导实验设计、实验实施、实验报告等整个动物实验研究过程规范性的内容清单和详解说明。《实验动物与比较医学》于2023年第43卷第2期发表了《〈动物研究：体内实验报告〉即ARRIVE 2.0指南的解释和阐述（一）》一文，基于国际期刊遵循ARRIVE 2.0指南的最佳实践，对该指南解读版原文前言和“关键10条”里的条目1~3进行了中文编译。本文将继续对“关键10条”里的条目4~7，包括“随机化”、“盲法”、“结果测量”和“统计方法”等部分进行相应编译介绍，以期促进国内研究人员充分理解并使用ARRIVE 2.0指南，提高动物实验研究及报告的规范性。

1.4 条目4：随机化

1.4.1 子条目4a：说明是否采取随机化方法将实验单元分配给对照组和实验组。如已随机化分配，需提供产生随机化序列的方法

详解：在进行分组时，使用适当的随机化分配方法能够确保各实验单元接受到特定处理的机会相同，且使每组实验单元的数量均衡。仅仅从笼盒里“随机”（即随意或任意）选择一只动物，不是真正统计学意义上的随机，因为它涉及人为判断。这种做法会带来偏倚并影响实验结果，原因在于研究人员进行动物分组时可能会有意或无意地做出一些主观判断，也可能由于不同实验组间的实验条件或动物本身就存在一些未知和不可控的差异。所以，使用经过验证的随机化方法有助于将选择性偏倚最小化，并减少分配到各组的动物在特征上存在系统化差异^[62-64]。由于非随机性分组得到的统计推断是无效的^[65-66]，因此对于旨在检验假设的任何实验，随机化都是先决条件。适当的随机化方法有多种，包括可以使用在线随机数生成器（例

如<https://www.graphpad.com/quickcalcs/randomize1/>），或使用Excel、Google Sheets、LibreOffice等电子表格软件中诸如“Rand（）”这样的函数。实验设计助手（Experimental Design Assistant，EDA）系统也具有随机化和分配隐藏的专用功能^[19]。

已有系统综述表明，动物实验如没有报告随机化或其他可减少偏倚的措施（例如盲法），更有报告夸大效应的可能，且这些效应符合传统的统计学显著性^[67-69]。在实验的全部或部分不能适用盲法的情况下，使用随机化就显得尤为重要。即使采用了随机化方法，研究人员的偏好也可能会扭曲分配情况，这种情况可使用分配隐藏来避免（见条目5：盲法）。在样本量较小的研究中，简单的随机化可能会导致组间不平衡；这时鼓励研究人员采用一些随机化的策略来进行平衡，如配对随机化^[70-72]和区组化^[17]。准确报告动物或实验单元分配至各组的方法，可使读者能够评估结果的可靠性，并识别潜在的局限性。

应报告使用的随机化类型（简单随机化、分层随机化、随机完全区组等），用于生成随机化序列的方法（例如计算机生成的随机化序列，以及所使用的算法或程序的详细信息），以及随机化的内容（例如对实验单元的处理，每只动物的处理顺序）。如果采用的随机化类型因实验而异，应报告每个实验的具体信息。如果将实验单元分配至各组时未采取随机化的方法，应明确说明并解释如何产生用来进行比较的各个实验组。

随机化策略的考量因素：（1）简单随机化。不考虑任何其他变量，将所有动物/样本同时随机分配到各实验组。这种策略很少使用，因为它不能确保其他可能影响实验结果的变量在比较组间是均衡的。（2）区组随机化。区组化法是一种控制实验单元之间自然变异的方法。该方法将一个实验分成更小的子实验（区组），并在每个区组内将实验处理随机分配给实验单元^[17,66,73]。这样就考虑到了可能使结果产生偏倚的干扰变量（例如笼盒位置、操作日期或周数）。分层随机

化采用与区组随机化相同的原理,只是分层往往是基于可能与反应相关的动物特征(例如体重等级或肿瘤大小等级)。与区组随机化相比,分层随机化在实际应用时可能出现一些差异(例如,在每个体重等级中可能不具有相等数量的实验单元)。(3)其他随机化策略。最小化是一种将动物/样本分配给实验组的替代策略,用来平衡可能影响实验结果的变量。在最小化的情况下,分配给下一只动物/样本的实验操作取决于已经被分配的那些动物/样本的特征。其目的在于每次分配都应尽量减少多个因素之间的不平衡^[74]。这种方法适用于连续性干扰变量,例如体重或起始肿瘤体积。

随机化策略中可考虑的干扰变量示例:(1)实验时间或日期;(2)窝、笼、鱼缸;(3)研究人员或手术人员(实施实验处理、进行手术或评估结果的人员,其经验和水平的不同会导致动物的应激水平或麻醉持续时间不同);(4)设备(例如PCR仪、分光光度计等,其校准可能会有所不同);(5)研究参数的测量(例如初始肿瘤体积);(6)动物特征(如性别、年龄范围、体重范围);(7)位置(在不同高度或不同笼架上的笼盒,其光照、通风和受到的干扰情况有所不同,这些可能会影响动物的重要生理过程)。

如果在随机化时使用了区组化因素,那么区组化因素也应该包括在分析中。干扰变量增加了样本的变异性,从而降低了统计效力。在分析中纳入一个干扰变量作为区组化因素,既可以解释这种变异性,也能增加统计的效力,从而可提高利用更少的实验单元检测真实效果的能力。但是,如果干扰变量对变异性没有实质性影响,区组化会耗损自由度,并因此降低效力。

报告示例1:“50只12周龄的SD大鼠,体重320~360 g,购自广东省医学实验动物中心(中国广州)。大鼠随机分为2组(25只/组):完整组和去势组。动物分组的随机数由微软Excel的Rand()函数生成”^[75]。

报告示例2:“初始慢性缺血/再灌注损伤后存活的动物进行随机分组。分组采用基于计算机的随机数字生成器”^[76]。

报告示例3:“在每个研究所,定期收集年龄匹配、具有相同遗传背景的野生型小鼠的两种性别的表型数据。每个管线、每种性别生成至少7只纯合子小鼠的队列……。小鼠随机分配到实验组(野生型与基因敲除型)是由孟德尔遗传法则驱动”^[29]。

1.4.2子条目4b:描述用于将潜在混杂因素最小化的策略,如处理和测量的顺序,或者动物/笼具的位置等。如果没有控制混杂因素,则需明确声明

详解:确保不同组动物之间除了实验操作外没有系统性差异,这是整个实验过程的一个重要原则。在实验设计和分析中,识别干扰变量(可能对结果产生潜在偏倚的变异源或条件)并对其进行处理,可以提高实验的灵敏度。例如,笼架顶部笼盒里的啮齿类动物可能会暴露在更高的光照水平下,这会影响动物的应激反应^[77]。

为了最大程度地减少实验过程中处理组之间出现的潜在差异,报告所实施的策略可以使他人能够评估实验的内部有效性。可报告的策略包括标准化(保持条件相同,例如由同一手术人员进行所有手术)、随机化(例如采样或测量顺序)、区组化或均衡化(例如动物笼或动物箱在架子上的位置)等,以确保各实验组受变异源的影响大致相同。在某些情况下,由于实际条件的限制,一些干扰变量的随机化受到阻碍,但它们仍然可以在分析中加以描述(见条目7:统计方法)。

在说明动物分组的方法时,应同时报告最小化混杂因素的方法。如果没有采取任何措施来最大限度地减少混杂因素(例如实验处理顺序、测量顺序、动物笼盒或动物箱在架子上的位置),请明确说明并解释原因。

报告示例1:“随机分组按照如下方法进行。从El-Nile公司抵达后,动物被指定组别并称其体重。一共有32只动物,分成4个不同体重组(每组8只动物)。各体重组别内的每一动物都被赋予一个临时随机数。根据笼盒在笼架上的位置,每个笼盒也被赋予一个数字编号。从每组所有的笼盒中随机选择一笼。从每个体重组中取2只动物,并在笼盒中给予它们永久的数字编号。然后,将笼盒随机分配至实验(暴露)组”^[78]。

报告示例2:“……测试时间为每天上午8:30至中午12:30,每天测试顺序随机。每只动物在每个测试日的不同时间被测试”^[79]。

报告示例3:“所有公牛按照体重分为4组,每组905只,每组内的公牛体重相似。按照完全随机区组设计,每组中的公牛被随机分配到4种实验处理中的一种,即每种实验处理905只公牛。动物被分配到20个围栏中(每个围栏181只,每种实验处理5个围栏)”^[80]。

1.5 条目5: 盲法

1.5.1 描述谁会在实验的不同阶段（动物分配、实验实施、结果评估、数据分析）知晓分组情况

详解: 研究人员通常期望特定的实验结果, 这可能会导致在无意中影响实验, 或以这种方式来解释数据以支持自己的研究假设^[81]。盲法就是一种用来尽量减少这些主观偏倚的策略。

尽管在直接比较盲法与非盲法评估结果的临床文献中, 有重要证据表明盲法的影响^[82], 但在动物研究中关于盲法影响的经验证据还比较有限^[83-84]。不过有来自系统综述的令人信服的数据表明, 非盲法结果评估会导致处理效果被高估, 而且缺乏随机化和盲法等减少偏倚的措施会使得效应量被夸大 30%~45%^[67-68, 85]。

理想情况下, 从实验开始到数据分析结束, 研究人员不应知晓动物已经或将要接受的处理。如果不能在实验的每个阶段都做到, 那么应该至少在其中某些阶段实施盲法。这对实验的组织会有一些影响, 可能需要额外人员的帮助, 例如手术人员的介入、技术员为动物用注射器进行编码、同事为实验组编码以进行分析等等。另外, 有在线资源可用于实施分配隐藏和盲法^[19]。

实验不同阶段的盲法包括: (1) 分配时。分配隐藏是指在动物分组完成之前, 不告知实施分组的人员将要分配给每只动物的实验处理。结合随机化, 分配隐藏有助于最大程度地减少选择性偏倚, 因为这种偏倚可能会在实验组间引入系统性差异。(2) 实验实施过程中。如果可能, 动物护理人员和进行实验处理的人员应不知晓分组情况, 以确保所有动物以相同的方式被看护、监测和处理。依据动物所接受的实验处理的不同而区别对待, 可能会改变动物的行为和生理并产生混杂。出于福利或安全原因, 有时不宜对动物护理人员使用盲法, 但在大多数情况下, 盲法是可行的。例如, 如果使用了有害微生物, 对照组动物可被视为和受感染的动物一样危险。如果某一福利问题出现于接受处理的动物时可短期容忍, 但出现于对照动物时不可接受, 就需要进行利弊分析, 以决定是否应该使用盲法。(3) 结果评估时。收集实验测量结果或进行评估的人员不应知道样本/动物接受了哪些处理, 以及哪些样本/动物被分组在一起。在结果评估过程中, 盲法尤为重要, 特别是在有主观因素的情况下 (例如, 在评估行为变化或组织学切片阅片时)^[83]。随机化测

试顺序也可以减少偏倚。如果无法对结果评估人员隐藏分组情况 (例如组间具有明显的表型或行为差异), 可将数据发送给第三方进行分析。第三方须对实验没有既得利益, 且不知道实验处理的预期是会改善还是恶化结果。通过以上这种方法可以减轻部分但不是全部的偏倚来源。(4) 数据分析时。数据分析人员应知晓哪些数据被分在一组, 以便进行分组比较, 但不应知晓每组接受了哪种特定的实验处理。这种盲法类型很重要, 但通常被忽视, 因为分析人员会做出许多半主观性决策, 例如将测量结果进行数据转换、选择处理缺失数据的方法以及处理异常值。如何做出这些决策也应该事前确定。

也可以在分析之前对数据进行编码处理, 从而在分析完成之前无法识别实验处理组。应具体说明实验过程的每个步骤或阶段是否使用了盲法, 并说明研究人员知晓或不知晓哪些特定实验处理或条件。如果在任何实验阶段中都没有使用盲法, 应明确说明这一点, 并给出不能或不考虑盲法的原因。

报告示例1: “对于每只动物, 有4名不同研究人员的参与, 情况如下: 第一名研究人员根据随机化表进行操作。该研究者是唯一知道实验处理组分配的人。第二名研究人员负责麻醉程序。第三名研究人员负责手术程序。最后, 由不知道分组和实验处理情况的第四名研究者来测量和评估实验结果”^[46]。

报告示例2: “……由于给药后会出现明显的行为性癫痫活动, 无法对实验者隐藏动物被注射的是毛果芸香碱还是生理盐水”^[86]。

报告示例3: “由于毛色的差异, 无法对研究人员隐藏小鼠品系。但是, 进行三室社交测试时, 使用头顶摄像机系统, 由ANY迷宫视频跟踪软件进行自动化的动物行为测试, 并进行无偏倚的数据分析。单室社交测试需要手动评分, 并由一位不了解实验目的的人员进行分析”^[87]。

1.6 条目6: 结果测量 (又称结局变量)

1.6.1 子条目6a: 清晰地定义所有评估的结果测量指标 (如细胞死亡、分子标志物或行为变化)

详解: 结局变量 (也称为因变量或响应变量) 是研究期间记录的任何变量 (例如受损组织的体积、死细胞的数量、特定分子标志), 以评估实验处理或实验干预的效果。结局变量对于表征样本 (如基线数据) 或描述复杂的反应 (例如“血流动力学”结局变量, 包括心率、血压、中心静脉压和心输出量) 可能很重

要。未能披露所有结局变量会导致在文献中引入偏倚，因为研究者往往更多地报告阳性结果（例如那些具有统计学意义的结果）^[88-91]。

明确阐述所测量的指标至关重要，尤其是当该指标可以通过不同的方式进行测量时。例如，动物的活动情况可以记录为移动时间或移动距离。在可能的情况下，结局变量应无偏倚地记录（例如对每个实验组的处理分配使用盲法。参见条目5：盲法）。需详细说明评估的结局变量与研究目标的相关性。

报告示例1：“评估了以下参数：阈值压力（TP，排尿前瞬时膀胱内压力）；排尿后压力（PVP，排尿后瞬时膀胱内压力）；峰值压力（PP，排尿时的膀胱内最高压力）；容量（CP，诱导第一次排尿所需的生理盐水体积）；顺应性（CO，CP与TP之比）；排尿收缩频率（VC）和非排尿收缩频率（NVCs）”^[92]。

1.6.2 子条目6b：对于检验假说的研究，明确主要的结果测量指标，即用于确定样本量的结果测量

详解：在假设验证实验中，主要结局变量用于回答主要的生物学问题。它是最重要的结果，在实验的计划阶段就需确定，并作为样本量计算的依据。而对于探索性研究，不必确定唯一的主要结局变量，通常会对多个实验结果进行评估（见条目13：研究目标）。

在一项旨在检测主要结局变量效应的假设检验研究中，次要结果的数据用于评估实验干预措施的其他效应，但次要结局变量的后续统计分析可能效力不足，这会使结果和解释变得不太可靠^[88,93]。声称检验假设但没有预先定义主要结局变量的实验，或在收集数据后改变了主要结局变量的实验（也称为主要结果转换），倾向于选择性地只报告具有统计学意义的结果，有利于阳性结果的发现^[94]。

提前注册实验方案可以保护研究人员免受选择性报告研究结果（也称为数据挖掘或P值操纵）的担忧，并为证明研究论文中报告的主要结果与研究计划的一致性^[95]提供证据（见条目19：实验方案注册）。

在使用推论统计来检验假设的研究中（例如t检验、方差分析），如果评估了一个以上的结果，要明确主要的结局变量，说明其是否在数据收集之前已明确定义，以及是否用于样本量计算。如果没有主要结局变量，请明确说明。

报告示例1：“本研究的主要结果是通过阶梯测试评估的前肢功能。次要结果包括旋转杆测试表现、脑卒中体积（利用MR成像或脑切片进行定量）、弥散张

量成像（DTI）连接组图和组织学分析，以评估神经元和小胶质细胞密度以及吞噬细胞活性等。该研究设计具有80%的检验效力，以检测阶梯试验中“抓到糖果颗粒”这一表现的相对25%的差异”^[96]。

报告示例2：“本研究的主要终点定义为随访结束时通过磁共振成像（MRI）评估的左心室射血分数（EF）。次要终点是通过MRI评估的左心室舒张末期容积（EDV）和左心室收缩末期容积（ESV），通过三苯基氯化四氮唑（TTC）共孵育后进行离体大体检查和延迟钆迟增强（LGE）MRI评估的梗死面积，通过压力容积（PV）-环和心脏超声检查连续评估的功能性参数，通过冠状动脉内压力和血流量评估的冠状动脉微血管功能，以及通过组织学评估的血管密度和纤维化”^[76]。

1.7 条目7：统计方法

1.7.1 子条目7a：提供用于每次分析的统计方法的细节，包括使用的软件

详解：实验采用的统计分析方法会反映实验的目的和设计。在数据收集前就需要确定统计分析方法（详见条目19：实验方案注册）。探索性研究和假设检验研究都可能使用描述性统计来总结数据（例如平均值和标准差，或中位数和极差）。在没有检验特定假设的探索性研究中，报告描述性统计数据对于产生新的假设很重要，这些新假设可能在后续实验中进行验证，但它不允许得出超出数据范围之外的结论。除了描述性统计之外，假设检验研究还可以使用推论统计来检验特定的假设。

详细报告分析方法对于确保读者和同行审稿人能够评估所选方法的适当性并判断分析结果的有效性至关重要。统计分析方法的描述应提供足够的细节，以便其他研究人员可以使用相同的方法重新分析原始数据并获得相同的结果。应明确哪种方法用于哪个分析。使用不同的方法分析数据，并有选择地报告那些具有统计学显著性的数据，会构成P值操纵（P-hacking，指为满足特定统计学标准而操纵数据）并在文献中引入偏倚^[90,94]。应完整地报告所有已执行的分析。描述统计方法的相关信息包括：（1）结局变量；（2）目标自变量；（3）在每项统计检验中所考虑的干扰变量（例如作为区组因素或协变量）；（4）所使用的统计分析和参考文献；（5）如何处理缺失值；（6）多重比较的调整；（7）所使用的软件和版本，如可能，应包括计算机代码^[97]。

结局变量可能受到所测试的处理或干预措施的影响,同时也可能受到其他因素的影响,如生物样本的特征(性别、窝、年龄、体重等)和技术因素(笼具、一天中的时刻、批次、实验者等)。为了降低偏倚风险,在实验设计中可以考虑其中的一些因素,例如在随机化中使用区组因素(详见条目4:随机化)。影响结果测量变异性的因素也应在分析中加以处理,例如作为区组因素(如试剂的批次或实验者)或作为协变量(如在随机化时的起始肿瘤大小)。

此外,为了正确地分析实验结果,认识到实验中可能存在的层次结构非常重要。层次结构可以引起集群效应。例如,当来自同一笼/窝的动物或来自同一动物的细胞,其测量结果彼此更相似时,可能会出现笼、窝或动物效应。这种关系必须在统计分析中进行处理,方法是在模型中纳入笼/窝/动物效应,或将结局变量汇总到笼/窝/动物水平。因此,描述实验的真实情况和数据的层次结构,以及为此而在设计和分析中所采取的措施,对于评估所使用的统计方法是否合适至关重要。

对于定制分析,例如有许多项的回归分析,详细描述分析路径是必要的。这可能包括详细描述初始模型和任何的模型简化步骤。

在报告描述性统计量时,应明确指出报告的是哪一种集中趋势指标(如平均值或中位数)和离散趋势指标(如标准差、极差、四分位数或四分位数极差)。同时描述在分析前对原始数据所做的任何修改(例如,相对于管家基因的目的基因表达相对量化)。有关统计报告的进一步指南,请参阅《已发表文献中的统计分析和方法(SAMPL)指南》^[98]。

报告示例 1:“使用 SAS (SAS Inst., Cary, NC) 的 GLM 程序进行方差分析。以猪圈作为实验单元,猪的生产性能参数记为猪圈平均值。该模型考虑了区组和饲料处置(5 种饲料)的影响。以初始体重为协变量对数据进行调整。用正交对比来检验 SDPP 处理(经紫外线处理 *vs* 未经紫外线处理)和饲料 SDPP 水平(3% *vs* 6%)的影响。结果以最小二乘均值表示。显著性水平设定为 $P < 0.05$ ”^[99]。

报告示例 2:“在单一模型中对所有感兴趣的风险因素进行了研究。逻辑回归考虑了区组因素,可以在控制其他变量的基础上明确研究每个自变量的效应……因为我们对饲养和环境影响感兴趣,所以在分析中通过对重要的生物学变量(年龄、背景品系、近交、性别、繁殖状况)进行区组化来控制它们的影响——

这些生物学变量在理发行为中的作用在其他地方有所描述,特别是将“理发”作为人类拔毛癖的模型中……本研究还通过按房间进行区组化来控制与此设计变量相关的未知环境变量的影响。检测了以下饲养和环境风险因素的效应:同笼动物的关系(如兄弟姐妹、非兄弟姐妹或混合),笼盒类型(如塑料或钢制),笼盒距地面的高度,笼盒的水平位置(笼盒是在笼架的侧面还是中间),饲养密度,以及笼盒里成年动物的数量。研究了笼盒离地板的高度与笼盒材料之间的相互作用,以及笼盒水平位置与笼盒材料之间的相互作用,然后从模型中删除了这几项,因为它们不显著。 $N = 1\,959$ 只小鼠被纳入该分析”^[100]。

1.7.2 子条目 7b: 描述用于评估数据是否能满足统计假设的任何方法,以及当统计假设无法满足时所做的方法变更

详解:假设检验通常是基于与基础数据相关的一些前提(条件)。应描述如何评估这些前提条件,以及数据是否满足这些前提,使读者能够评估所使用的统计方法的适用性。如果前提不正确,结论就可能不成立。例如,使用参数检验(如 t 检验、 z 检验、方差分析等)的前提是数据是连续变量,分析的残差符合正态分布,各组响应相互独立,且各组总体方差齐。

有多种正态性检验的方法,例如 Shapiro-Wilk 检验和 Kolmogorov-Smirnov 检验。然而,这些检验必须谨慎使用。如果样本量较小,它们将难以检测到非正态性;如果样本量较大,将检测到不重要的偏离。另一种方法是用可视化图形来评估数据,例如正态概率图、箱型图、散点图。如果分析的残差不符合正态分布,则通过数据转换有可能满足前提条件,即将相同的数学函数(如 $\log_e$ 、 $\log_{10}$ 、平方根)应用于所有数据以产生符合正态分布数据。

其他类型的结局变量(二分类、多分类或等级变量)则需要不同的分析方法,每种方法都有不同的前提条件。例如,分类数据应按计数和百分比或比例进行汇总,并通过比例检验进行分析;这些分析方法的前提是,相关数据是二分类、等级或名义变量,并且是独立的^[18]。

对于每一个统计检验(参数或非参数),需要报告结局变量的类型和用于检验统计学前提(是否得到满足)的方法。如果对数据进行了转换,应准确指出所使用的转换方法,以及应用于哪些结局变量。如果不满足前提条件,并且使用了替代方法(如使用了非参

数检验,它不需要有正态性),则需要报告分析中的任何改变。如果未检验数据是否满足前提条件,则应明确说明。

报告示例 1:“使用 Shapiro-Wilk 正态检验和 Levene's 方差齐性检验,并通过检查残差和拟合值图来检验模型假设(是否被满足)。一些响应变量必须采用自然对数或平方根或立方根进行转换,但为了显示显著效应而进行了反向转换”^[101]。

报告示例 2:“采用固定效应两因素方差分析(fixed-effects 2-way ANOVA)评估饲养条件(实验处理)和安乐死对皮质醇水平的影响。初步的探索性分析表明,平均皮质醇水平较高的组在这一反应变量上也有较大的差异。为了使差异更加均匀,我们在分析中使用了每条鱼单位体重的皮质醇的对数变换作为因变量。这使得分析前提即正态性和方差齐性(标准差相等)得以满足”^[102]。

(未完待续)

[作者贡献 Author Contribution]

陈国元负责 ARRIVE 2.0 指南完整版“关键 10 条”中第 4~5 条的编译初稿,以及本文内容的整合和修改;
卢晓负责 ARRIVE 2.0 指南完整版“关键 10 条”中第 6 条的编译初稿,以及本文内容的修改;
白玉负责 ARRIVE 2.0 指南完整版“关键 10 条”第 7 条的编译初稿,以及本文内容的修改;
于灵芝和乔颖分别参与 ARRIVE 2.0 指南完整版“关键 10 条”第 6 和第 7 条的编译初稿写作;
王剑、卢今、刘晓宇和卢选成负责对 ARRIVE 2.0 指南完整版“关键 10 条”中第 4~7 条的编译内容提出关键性修改建议;
高静负责 ARRIVE 2.0 指南完整版中文编译内容中有关统计学概念及阐述的审定;
李焱负责 ARRIVE 2.0 指南完整版中文编译工作的策划、组织,以及提出关键性修改建议;
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]

- [62] SCHULZ K F, CHALMERS I, HAYES R J, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials[J]. JAMA, 1995, 273(5): 408-412. DOI: 10.1001/jama.273.5.408.
- [63] SCHULZ K F, GRIMES D A. Allocation concealment in randomised trials: defending against deciphering[J]. Lancet, 2002, 359(9306):614-618. DOI: 10.1016/S0140-6736(02)07750-4.
- [64] CHALMERS T C, CELANO P, SACKS H S, et al. Bias in treatment assignment in controlled clinical trials[J]. N Engl J Med, 1983, 309(22): 1358-1361. DOI: 10.1056/NEJM.1983.12013092204.
- [65] GREENBERG B G. Why randomize?[J]. Biometrics, 1951, 7(4): 309. DOI: 10.2307/3001653.
- [66] ALTMAN D G, BLAND J M. Statistics notes. Treatment allocation in controlled trials: why randomise?[J]. BMJ, 1999, 318(7192):1209. DOI: 10.1136/bmj.318.7192.1209.
- [67] HIRST J A, HOWICK J, ARONSON J K, et al. The need for randomization in animal trials: an overview of systematic reviews[J]. PLoS One, 2014, 9(6): e98856. DOI: 10.1371/journal.pone.0098856.
- [68] VESTERINEN H M, SENA E S, FFRENCH-CONSTANT C, et al. Improving the translational hit of experimental treatments in multiple sclerosis[J]. Mult Scler, 2010, 16(9): 1044-1055. DOI: 10.1177/1352458510379612.
- [69] BEBARTA V, LUYTEN D, HEARD K. Emergency medicine animal research: does use of randomization and blinding affect the results?[J]. Acad Emerg Med, 2003, 10(6):684-687. DOI: 10.1111/j.1553-2712.2003.tb00056.x.
- [70] TAVES D R. Minimization: a new method of assigning patients to treatment and control groups[J]. Clin Pharmacol Ther, 1974, 15(5):443-453. DOI: 10.1002/cpt1974155443.
- [71] SAINT-MONT U. Randomization does not help much, comparability does[J]. PLoS One, 2015, 10(7): e0132102. DOI: 10.1371/journal.pone.0132102.
- [72] LAAJALA T D, JUMPPANEN M, HUHTANIEMI R, et al. Optimized design and analysis of preclinical intervention studies *in vivo*[J]. Sci Rep, 2016, 6: 30723. DOI: 10.1038/srep30723.
- [73] KANG M, RAGAN B G, PARK J H. Issues in outcomes research: an overview of randomization techniques for clinical trials[J]. J Athl Train, 2008, 43(2):215-221. DOI: 10.4085/1062-6050-43.2.215.
- [74] ALTMAN D G, BLAND J M. Treatment allocation by minimisation[J]. BMJ, 2005, 330(7495): 843. DOI: 10.1136/bmj.330.7495.843.
- [75] ZHAO S K, KANG R, DENG T, et al. Comparison of two cannulation methods for assessment of intracavernosal pressure in a rat model[J]. PLoS One, 2018, 13(2): e0193543. DOI: 10.1371/journal.pone.0193543.
- [76] JANSEN OF LORKEERS S J, GHOSH J M I H, KOUDESTAAL S, et al. Xenotransplantation of human cardiomyocyte progenitor cells does not improve cardiac function in a *Porcine* model of chronic ischemic heart failure. results from a randomized, blinded, placebo controlled trial[J]. PLoS One, 2015, 10(12): e0143953. DOI: 10.1371/journal.pone.0143953.
- [77] ISHIDA A, MUTOH T, UYAMA T, et al. Light activates the adrenal gland: timing of gene expression and glucocorticoid release[J]. Cell Metab, 2005, 2(5): 297-307. DOI: 10.1016/j.cmet.2005.09.009.
- [78] EL-AGROUDY N N, EL-NAGA R N, EL-RAZEQ R A, et al. Forskolin, a hedgehog signalling inhibitor, attenuates carbon tetrachloride-induced liver fibrosis in rats[J]. Br J Pharmacol, 2016, 173(22):3248-3260. DOI: 10.1111/bph.13611.
- [79] CARRILLO M, MIGLIORATI F, BRULS R, et al. Repeated

- witnessing of conspecifics in pain: effects on emotional contagion[J]. PLoS One, 2015, 10(9): e0136979. DOI: 10.1371/journal.pone.0136979.
- [80] DEL BIANCO BENEDETI P, PAULINO P V R, MARCONDES M I, et al. Partial replacement of ground corn with glycerol in beef cattle diets: intake, digestibility, performance, and carcass characteristics[J]. PLoS One, 2016, 11(1): e0148224. DOI: 10.1371/journal.pone.0148224.
- [81] NUZZO R. How scientists fool themselves - and how they can stop[J]. Nature, 2015, 526(7572):182-185. DOI: 10.1038/526182a.
- [82] HRÓBJARTSSON A, THOMSEN A S S, EMANUELSSON F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors[J]. BMJ, 2012, 344: e1119. DOI: 10.1136/bmj.e1119.
- [83] ROSENTHAL R, FODE K L. The effect of experimenter bias on the performance of the albino rat[J]. Syst Res, 2007, 8(3):183-189. DOI: 10.1002/bs.3830080302.
- [84] ROSENTHAL R, LAWSON R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats[J]. J Psychiatr Res, 1964, 2(2): 61-72. DOI: 10.1016/0022-3956(64)90003-2.
- [85] MACLEOD M R, VAN DER WORP H B, SENA E S, et al. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality[J]. Stroke, 2008, 39(10):2824-2829. DOI: 10.1161/STROKEAHA.108.515957.
- [86] NEUMANN A M, ABELE J, KIRSCHSTEIN T, et al. Mycophenolate mofetil prevents the delayed T cell response after pilocarpine-induced status epilepticus in mice[J]. PLoS One, 2017, 12(11): e0187330. DOI: 10.1371/journal.pone.0187330.
- [87] HSIEH L S, WEN J H, MIYARES L, et al. Outbred CD1 mice are as suitable as inbred C57BL/6J mice in performing social tasks[J]. Neurosci Lett, 2017, 637: 142-147. DOI: 10.1016/j.neulet.2016.11.035.
- [88] JOHN L K, LOEWENSTEIN G, PRELEC D. Measuring the prevalence of questionable research practices with incentives for truth telling[J]. Psychol Sci, 2012, 23(5):524-532. DOI: 10.1177/0956797611430953.
- [89] DWAN K, ALTMAN D G, ARNAIZ J A, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias[J]. PLoS One, 2008, 3(8): e3081. DOI: 10.1371/journal.pone.0003081.
- [90] TSILIDIS K K, PANAGIOTOU O A, SENA E S, et al. Evaluation of excess significance bias in animal studies of neurological diseases[J]. PLoS Biol, 2013, 11(7): e1001609. DOI: 10.1371/journal.pbio.1001609.
- [91] SENA E S, VAN DER WORP H B, BATH P M W, et al. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy[J]. PLoS Biol, 2010, 8(3): e1000344. DOI: 10.1371/journal.pbio.1000344.
- [92] CLAUDINO M A, LEIRIA L O S, DA SILVA F H, et al. Urinary bladder dysfunction in transgenic sickle cell disease mice[J]. PLoS One, 2015, 10(8): e0133996. DOI: 10.1371/journal.pone.0133996.
- [93] LANDIS S C, AMARA S G, ASADULLAH K, et al. A call for transparent reporting to optimize the predictive value of preclinical research[J]. Nature, 2012, 490(7419): 187-191. DOI: 10.1038/nature11556.
- [94] HEAD M L, HOLMAN L, LANFEAR R, et al. The extent and consequences of p-hacking in science[J]. PLoS Biol, 2015, 13(3): e1002106. DOI: 10.1371/journal.pbio.1002106.
- [95] MUNAFÒ M R, NOSEK B A, BISHOP D V M, et al. A manifesto for reproducible science[J]. Nat Hum Behav, 2017, 1:0021. DOI: 10.1038/s41562-016-0021.
- [96] EMMRICH J V, NEHER J J, BOEHM-STURM P, et al. Stage 1 Registered Report: effect of deficient phagocytosis on neuronal survival and neurological outcome after temporary middle cerebral artery occlusion (tMCAo)[J]. F1000Research, 2017, 6:1827. DOI: 10.12688/f1000research.12537.3.
- [97] BRITISH ECOLOGICAL SOCIETY. A guide to reproducible code in ecology and evolution[S/OL]. 2017 [2020-06-16]. <https://www.britishecologicalsociety.org/wp-content/uploads/2017/12/guide-to-reproducible-code.pdf>.
- [98] LANG T A, ALTMAN D G. Basic statistical reporting for articles published in biomedical journals: the Statistical Analyses and Methods in the Published Literature or the SAMPL Guidelines[J]. Int J Nurs Stud, 2015, 52(1): 5-9. DOI: 10.1016/j.ijnurstu.2014.09.006.
- [99] POLO J, RODRÍGUEZ C, RÓDENAS J, et al. Ultraviolet light (UV) inactivation of *Porcine parvovirus* in liquid plasma and effect of UV irradiated spray dried *Porcine* plasma on performance of weaned pigs[J]. PLoS One, 2015, 10(7): e0133008. DOI: 10.1371/journal.pone.0133008.
- [100] GARNER J P, DUFOUR B, GREGG L E, et al. Social and husbandry factors affecting the prevalence and severity of barbering ('whisker trimming') by laboratory mice[J]. Appl Anim Behav Sci, 2004, 89(3-4): 263-282. DOI: 10.1016/j.applanim.2004.07.004.
- [101] NEMETH M, MILLES E, WAGNER K H, et al. Sex-specific effects of diets high in unsaturated fatty acids on spatial learning and memory in Guinea pigs[J]. PLoS One, 2015, 10(10): e0140485. DOI: 10.1371/journal.pone.0140485.
- [102] KECK V A, EDGERTON D S, HAJIZADEH S, et al. Effects of habitat complexity on pair-housed zebrafish[J]. J Am Assoc Lab Anim Sci, 2015, 54(4):378-383.

(收稿日期: 2023-03-30 修回日期: 2023-06-18)

(本文编辑: 张俊彦, 富群华)

引用本文

陈国元, 卢晓, 白玉, 等. 《动物研究: 体内实验报告》即 ARRIVE 2.0 指南的解释和阐述 (二) [J]. 实验动物与比较医学, 2023, 43(3): 324-333. DOI: 10.12300/j.issn.1674-5817.2023.042.

CHEN G Y, LU X, BAI Y, et al. Explanation and elaboration of the ARRIVE guidelines 2.0—reporting animal research and *in vivo* experiments (II) [J]. Lab Anim Comp Med, 2023, 43(3): 323-331. DOI: 10.12300/j.issn.1674-5817.2023.042.