

曹策,李玲美,訾明杰,等. 医学研究中动物实验样本量的确定方法 [J]. 中国比较医学杂志, 2023, 33(2): 99-105.
Cao C, Li LM, Zi MJ, et al. Methods for determining animal sample sizes in medical experimental studies [J]. Chin J Comp Med, 2023, 33(2): 99-105.
doi: 10.3969/j.issn.1671-7856.2023.02.013

医学研究中动物实验样本量的确定方法

曹 策^{1,2}, 李玲美², 訾明杰², 付建华², 任钧国², 李 磊^{2*}, 刘建勋^{1,2*}

(1. 广东药科大学中医药研究院, 广州 510006; 2. 中国中医科学院西苑医院 中药药理北京市重点实验室, 国家中医心血管病临床医学研究中心, 北京 100091)

【摘要】 在开始人体临床药物试验之前, 研究人员需要在动物实验中进行广泛的临床前研究, 以初步明确药物的作用、毒性和药代动力学特征。在实验研究中, 除了与研究设计有关的一般考虑因素外, 动物样本量计算也是核心要素。目前研究人员在动物样本量的选择方面具有很大的随意性, 不仅导致实验结果的可复制性和代表性降低, 而且造成了大量资金的浪费。因此, 本文将医学实验研究分为探索性研究和验证性研究, 并且探讨了基于文献与经验的样本量确定法、精度分析确定法、两样本均数确定法、资源方程式确定法和 KISS 确定法, 旨在为实验中样本量的选择提供计算方案, 进而降低对动物伤害与提高实验的准确性。随着医学统计学与医学实验结合的发展和传播, 动物实验样本量确定方面的研究将愈加完善, 将为研究过程的严谨性及研究结论的可靠性提供坚实的理论基础和实验依据。

【关键词】 实验动物; 样本量; KISS 分析; 精度分析

【中图分类号】 R-33 **【文献标识码】** A **【文章编号】** 1671-7856 (2023) 02-0099-07

Methods for determining animal sample sizes in medical experimental studies

CAO Ce^{1,2}, LI Lingmei², ZI Mingjie², FU Jianhua², REN Junguo², LI Lei^{2*}, LIU Jianxun^{1,2*}

(1. Institute of Chinese Medicine Sciences, Guangdong Pharmaceutical University, Guangzhou 510006, China.

2. Institute of Basic Medical Sciences of Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing Key Laboratory of Chinese Materia Pharmacology, National Clinical Research Center of Traditional Chinese Medicine for Cardiovascular Diseases, Beijing 100091)

【Abstract】 Before beginning human clinical drug trials, researchers need to conduct extensive preclinical studies involving animal experiments to preliminarily clarify each drug's effects, toxicity, and pharmacokinetics. In addition to general considerations related to study design, the calculations for determining the appropriate sample sizes are a core element of experimental studies. Currently, there is a great deal of randomness in the selection of animal sample sizes, which not only leads to a reduction in the reproducibility and representativeness of experimental result but also causes funds to be wasted. Therefore, in this paper, we divide medical experimental research into exploratory and confirmatory studies

【基金项目】 国家自然科学基金项目(82174015); 国家中医药管理局中医药传承与创新“百千万”人才工程(岐黄工程)岐黄学者; 中国中医科学院科技创新工程(CI2021A04609)。

【作者简介】 曹策(1996—), 男, 在读硕士研究生, 研究方向: 中西医结合基础。E-mail: caoce1@foxmail.com

【通信作者】 李磊(1982—), 男, 博士, 副研究员, 硕士生导师, 研究方向: 心血管中药药理研究及新药研究与开发。

E-mail: lilei0502@126.com

刘建勋(1955—), 男, 博士, 研究员, 博士生导师, 研究方向: 中药药理及新药研究与开发研究。E-mail: liujx0324@sina.com

* 共同通信作者

and provide examples of sample size determination method based on the literature and our experience. Precision analysis, two-sample mean, resource equation, and KISS approach determination method are described with the aim of providing a calculation scheme for the selection of sample sizes and, therefore, reduce unnecessary harm to animals and improving the accuracy of the experiments. With the development and dissemination of combinations of medical statistics and medical experiments, research on the determination of animal experimental sample sizes will become increasingly perfected, providing firm theoretical and experimental bases for a rigorous research process and reliable conclusion.

【Keywords】 laboratory animal; sample size; KISS analysis; precision analysis

Conflicts of Interest: The authors declare no conflict of interest.

随着我国对生物医药产业的大力支持,全行业研发投入大幅增长,医学实验动物作为生物医药行业发展重要的支撑条件之一,在疾病致病机理研究、药品研发、现代中药、疫苗和诊断试剂、医疗设备与器材等诸多领域都起着不可或缺的作用,需求量不断增长。合理有效的使用实验动物是研究结论科学性和可靠性的保证,也是医学实验动物伦理学原则的要求。合理应用实验动物的诸多考虑因素中,实验动物的样本含量(简称样本量)大小确定无疑是一个核心要素。

样本量是指动物实验所需要的动物数量,其大小直接关乎研究的重复性,如果样本量太小,研究结果的可重复性及代表性就欠佳可能得出假阴性或假阳性的结论。近年来,动物实验研究的可重复性问题越来越引起科学家、基金资助方和政策制定者的极大关注。2020 年 7 月在 *PLoS Biology* 杂志上发表了《关于动物实验研究报告的内容清单的 ARRIVE 指南 2.0 版》,主要是为了确保动物实验的研究报告中包含实验设计、动物基本信息、样本量等足够的研究细节,明确相关的知识内容,进而使读者和审稿人充分地审查研究内容,评估研究方法的严谨性,以重复研究的方法或结果^[1]。在实验结果的可重复性方面,众多论文的结果无法进行再次复刻,通过对动物论文的调查报告发现,大量的论文中没有说明动物的性别、年龄或体重,并且没有一篇论文阐述了选择实验动物样本量的理由,这可能是实验结果无法进行复制的原因之一^[2-3]。

然而,如果样本量过大,试验过程中对研究对象的潜在伤害越大,还存在一定伦理问题,而且研究所需的经费和资源就越多,项目执行的难度也越大。在实验动物的伦理方面,一方面,实验中每组的样本量应当尽可能的降低,但是同时也应该在合理范围内,以便获得足量的目标实验数据^[4]。另一方面,实验中的组数也应当尽可能缩减,避免不必要的组别设立。组数和组内样本量决定着动物实

验的整体最佳样本量。一般来说,通过正确、合理的计算来确定实验研究中的动物数量,可以避免实验中数量不足或数量过多的情况发生^[5],进而符合动物伦理中的“3R”原则。在经济效应方面,据统计,由于论文实验结果的不可复制性,仅仅美国一年在此方面便浪费了 280 亿美元^[6],这提示我们应当在动物研究开始前计算所需样本量,在达到检验预期目标的基础上,选择合适的样本量^[7]。过小的样本量会降低实验识别显著差异的灵敏度,而过大的样本量则会浪费时间、成本和资源^[8]。

样本量在实验结果、动物伦理和经济效应方面具有的突出、核心和重要作用,这要求我们在医学动物实验中必须提前计算出合适、准确的样本量。因此本文通过介绍实验研究的不同类型,随后介绍适合不同类型的动物样本量计算方法,并通过对相关案例的计算进一步阐述研究类型的选择和样本量计算的原理,以期后续实验研究中动物样本量的确定提供理论基础。

1 实验研究类型与样本量估算

确定合适的样本量是研究设计的关键步骤,然而不同实验设计类型的动物样本量估算方法不尽相同。医学实验研究通常分为探索性研究和验证性研究,如何合理确定探索性研究的样本量,以及选用何种精确度的参数用于验证性样本量的估算是研究人员在实验设计之初就应考虑的问题。

1.1 探索性研究

进行探索性研究是为了检查主要研究或关键研究中打算测量的变量的测量精度的可行性,并验证预实验的逻辑性质。此外,探索性研究也被用来创造新的假设。换句话说,这些研究的目的是确定某种反应的趋势或模式,因此,它并不需要进行显著性检验。例如,传统观念认为心肌细胞是高度分化的细胞,因此不具有增殖和再生的功能,但是随着科技的进步,传统观念逐渐被打破,科学家开始

逐步摸索心肌再生的通路和机制,因为这是以往并未涉及的全新方向,因此属于探索性研究。这些研究的样本量有时是根据以前的研究计算的,但是大多数情况下,并无以往的研究数据或者以往数据不可用,此时,探索研究的样本量要基于研究人员的经验或猜测而进行。通过从这些研究中获得的数据(标准差、两个样本之间的平均差等),用于计算正式研究的样本量^[9]。探索性研究常用的样本量计算方法有资源方程式确定法和 KISS 确定法。

1.2 验证性研究

验证性研究是一种对照研究,其中的研究假设是事先完成的,并且设计得较为完善。研究的目的就是完成对预先假定的内容的验证,通过实验的测试进而完成假设到结论的转化^[10]。例如,我们通过网络药理学推测某种药物具有抗炎的作用,但是为了证实我们的猜想,我们需要首先在动物实验中进行验证,此类研究均属于验证性研究。在验证性研究中,选取适当的精度来估计处理目标的影响大小是极为重要的,其中适当的样本量更是其中的关键环节,样本量过小不具有广泛意义,样本量过大会造成时间、精力和资金的浪费,样本量的计算在为验证性研究提供证据支持这一方面起着重要作用。因此,估计研究的有效样本量就显得尤为重要。验证性研究常用的样本量计算方法有基于文献与经验的样本量确定方法、精度分析确定法和两样本均数确定法等。

2 基于文献与经验的样本量确定方法

目前,对于大部分从事动物实验的研究人员而言,借鉴本研究方向的其他人员一贯使用的成果中的样本量是最常见的方法。因为研究方向的类似,所以在此条件下的自变量、因变量、实验用量、实验物种和结果变量均具有借鉴意义,不仅可以简化实验方案的设计,而且还可以保证基本成果的产出,降低了出现实验结果出现偏移的可能性。一般情况,小动物(小鼠、大鼠)每组 10~30 只,中等动物(兔、豚鼠等)每组 8~20 只,大动物(狗、猪、猫、猴等)每组 5~15 只。此外,对于如蛋白质组、转录组、Western blot 等实验成本较高、方法复杂的检测,一般样本量最低为 3 也是可以接受的。

3 基于统计学的样本量确定方法

3.1 样本量估算相关的统计学概念

总体(population),是根据研究目的确定的,所

有同质研究对象某一(组)指标值的集合,描述总体特征的指标称为参数。样本(sample),是从总体中随机抽取的、数量足够的,能代表总体特征的部分对象某一(组)指标的集合,描述样本特征的指标称为统计量。

医学科学研究的目的是由样本推断总体,依据统计量的特征或性质对总体的参数做出推断,进而阐明总体的特征与规律。根据样本统计量作出推断结论时,由于样本信息的局限性,其结论有可能会出现问题,通常可以发生两类错误,分别称为 I 型错误和 II 型错误,见表 1。I 型错误是把本来无统计学联系的错误地判断为有统计学联系,即犯假阳性错误。这类错误形成的概率,称为显著性水平(significance level),一般用 α 表示, α 越小则样本量需求越大。II 型错误是把本来有统计学联系的错误地判定为无统计学上的联系,即犯假阴性错误。这类错误形成的概率,称假阳性概率,用 β 表示,并把 $(1-\beta)$ 称为检验的效能(power)。效能的含义为,当统计量与总体实际上有统计学联系时,对样本含量为 n 的群体进行研究时,有 $(1-\beta)$ 的把握得出差别有显著意义的结论。 β 越小,检验效能越大,所需样本量越大。

标准差(s),主要反映数据的变异程度,其值越大,则需要的样本量越大。效应(effect size)一般指容许误差(δ),即两总体参数间的差别,容许误差越小,所需样本量越大。若研究者无法得到总体参数间差值(δ),可以通过预实验的组间差距作为容许误差,或根据专业知识确定有专业意义的差值作为容许误差,如平均舒张压的差值 >5 mmHg;当没有给定专业意义容许误差水平的情况下,用 0.25 倍或 0.50 倍的标准差来设定容许误差。

3.2 精度分析确定法

精度分析是计算样本量的方法之一。该方法在固定的置信度水平下,存在期望的精度,即固定的 I 型错误,进而计算样本大小^[11]。精度分析相比

表 1 可能发生的两类错误图
Table 1 Two types of errors that may occur

客观实际 Objective reality	假设检验的结果 Result of the hypothesis test	
	拒绝 H_0 Reject H_0	接受 H_0 Accept H_0
H_0 成立 H_0 established	I 型错误(α) Type I error	推断正确($1-\alpha$) Inference is correct
H_1 成立 H_1 established	推断正确($1-\beta$) Inference is correct	II 型错误(β) Type II error

较其他方案更为简单和便于计算,但可能有很小的概率发生真正的差异。预先给定的概率(1- α)所确定的包含未知总体参数的范围称为该参数的可信区间或者置信区间,预先给定的概率 1- α 称为可信度或者置信度,常取 95%或 99%。可信区间的精度取决于其区间的宽度,较窄的区间具有较精确的精度^[12]。当标准差 s 已知时,对于 1- α 可信区间所需的样本量计算公式为:

$$n = \frac{(Z_{\alpha/2})^2 s}{E^2}$$

其中, $Z_{\alpha/2}$ 是标准正态分布的第 $\alpha/2$ 个分位数, E 是总体均数(\bar{x})估计中的最大误差。

例如:在某一中药成分对小鼠缺血再灌注损伤后对炎症的影响中,实验时比较 A 处理组和 B 处理组的炎症因子增加量,设两处理动物的总体标准差为 1.2 pg/mL,最大误差不超过 0.30 pg/mL,若要求以 $\alpha=0.05, \beta=0.10$,那需要设置多少只实验动物?

已知双侧 $\alpha=0.05, Z_{0.05/2}=1.96; \beta=0.10; s=1.2; E=0.30$ 。代入公式可得:

$$n = \frac{(Z_{\alpha/2})^2 s}{E^2} = \frac{1.96^2 \times 1.2}{0.30^2} = 51.22 \approx 52。$$

因此,本实验共需设置 52 只小鼠,即 A 组和 B 组各 26 只。

3.3 两样本均数确定法

两样本均数确定法是在容许误差 δ 、总体标准差 s 、 α 和 β 已知的情况下,计算出实验动物样本量^[13]。在实际工作中,在未指定 δ 情况下,可对 δ/s 进行适当假设来估算样本含量 n ;在指定 δ 情况下,可以用样本标准差代替总体标准差。两样本均数比较所需样本含量可通过查表直接得到,或采用计算公式如下:

$$n_1 = n_2 = 2 \left[\frac{(\bar{x}_\alpha + \bar{x}_\beta)}{\delta/s} \right]^2 + \frac{1}{4} \bar{x}_\alpha^2$$

其中, n_1 和 n_2 分别代表两样本所需要的样本量; δ 为 $\bar{x}_\alpha - \bar{x}_\beta$,为两总体均数之差值; s 为总体标准差(假设两总体标准差相等); α 有单侧双侧之分, β 只取单侧。

例如:在作两种处理动物冠状动脉血流量实验时,比较 A 处理动物和 B 处理动物的平均血流量增加,设两处理组动物的标准差相等。若要求以 $\alpha=0.05, \beta=0.10$,达到能辨别出两者均数增加的差别是其标准差的 60%,那需要设置多少只实验动物?

已知双侧 $\alpha=0.05, \bar{x}_{0.05/2}=1.96; \beta=0.10, \bar{x}_{0.10}$

$=1.282; \delta/s=0.6$ 。代入公式可得:

$$n_1 = n_2 = 2 \left[\frac{(\bar{x}_\alpha + \bar{x}_\beta)}{\delta/s} \right]^2 + \frac{1}{4} \bar{x}_\alpha^2 = 2 \left[\frac{(1.96 + 1.282)}{0.6} \right]^2 + \frac{1}{4} 1.96^2 = 59.4 \approx 60。$$

因此,每组实验动物样本量为 60 只,总体实验样本量为 120 只。

3.4 资源方程式确定法

在动物研究中,特别是探索性的实验中,标准差和效应值在文献中可能不容易获得,或者很难估计,此时,资源方程式(resource equation)便是最好的计算方案,它在方差分析中设置误差自由度的可接受范围,这适用于结果是一个定量变量并且适合用方差分析的研究^[14]。资源方程式计算公式为:

$$E = N - K = Kn - K = K(n-1)$$

其中 N 为实验单位的总数, n 为每组的样本量, K 为处理组的数量, E 为两者之间的差值,上述公式可以转换如下: $n=E/K+1$ 。 E 的最小值为 10,最大值为 20。因此可以获得每组动物的最小和最大数量: $Min n=10/K+1, Max n=20/K+1$ 。因此,总体而言,所需的最小和最大动物数量为: $Min N=Min n \times K=(10/K+1) \times K, Max N=Max n \times K=(20/K+1) \times K$ 。

例如:在研究某药物对于大鼠的心肌梗死面积的影响,根据实验安排需要设计三组,那整个实验需要多少只大鼠?

已知 $K=3$,代入公式可得: $Min N=Min n \times K=(10/K+1) \times K=(10/3+1) \times 3=12.99 \approx 13, Max N=Max n \times K=(20/K+1) \times K=(20/3+1) \times 3=22.98 \approx 23$ 。

因为分组数 3 为奇数,因此略作修正后总体实验样本量最少为 15 只,最多为 24 只。

实验单位的总数针对不同的实验,其具体含义是随之改变的。在计划实验时,准确地确定实验单位是什么。一个实验中通常有多个可能的选择,如果不准确界定实验单位,随后对信息的分析和从结果得出的结论可能会产生误导^[15]。例如,当要比较不同的药物时,一个动物往往是自然的实验单位。每个动物接受一种药物,对药物反应的测量是在动物整体上进行的,或者是在血液或组织的样本上,亦或者是对动物的反应上进行测量,通过详细记录并进行分析,进而得出结论。在这个实验其中,实验单位就涉及是动物整体、血液或组织或者动物反应等 3 种可能。此外,对于所有实验单位,每次测量

都应当以相同的方式进行的,除了药物,还有其他形式的治疗,包括各种形式的手术、不同方法的物理治疗或者不同类型的饲养方式。通过本实验分析出的结论,将适用于本实验单位可以代表的一类群体。因此,明确实验单位是什么对于实验结果的确定和推广是极其重要的。

处理组的数量应当为除了空白组之外的所有组数,任何做了处理的组(包括模型组)都应包含在其中。由此计算出的 E 值,虽然没有确切的规定,但是 Mead R 认为 E 的值应该介于 10~20 之间。倘若 E 值过小,就可能会增加犯 II 型错误(β 错误)的可能,即当原假设 H_0 不正确(即 H_1 正确),但由于检验统计量的值没有落入拒绝域而未被拒绝时所犯的错误的。但是如果 E 值过大,也会造成实验中使用动物和资金的浪费。

3.5 KISS 确定法

KISS 确定法即通过预先假设的临时样本量,在标准差、检验效能、显著性水平、效应量(effect size, ES)、检验的单、双侧和样本量均确定的基础上,通过标准效应量查表,从而得到计算出的可检测效应量,最后将计算出的效应量与所需的效应量比较,如果希望够检测到更小的效应量,可以通过增加临时样本大小并重新进行计算。其中标准效应量(standardized effect size, SES)的计算公式如下: $SES = ES_r/s$,其中 ES_r 是计算出的被比较两组间平均值之差, s 是整体的标准差,其计算公式如下:

$$s = \sqrt{(s_1^2 + s_2^2)/2}$$

其中 s_1^2 和 s_2^2 为两组的标准差。因此, SES 实际上是两组平均值之间的差异的大小。根据表 2 可知,如果实验中的检验效能、侧度和显著性水平固定,则 SES 与样本量直接相关。

例如:当研究一种药物对小鼠的红细胞计数影响时,已知 C57 小鼠的平均红细胞计数为 9.19,整体标准差为 0.70(n/mL),则本实验需要的每组小鼠样本量为多少?

已知 $s=0.70$,检验效能为 90%,双侧检验,我们假设所需样本量为 12 只,由表 2 可知, $SES = 1.39$,即 $ES_r = SES \times s = 1.39 \times 0.70 = 0.97$, $P = (0.97/9.19) \times 100 = 11\%$ 。因此,12 只小鼠/组的样本量具有检测出 11%范围内变化的能力。

样本标准差要从先前的实验中获得,涉及动物的种类、品系、年龄、性别和健康状况等,实验需严格控制个体间的差异。 ES 是被比较两组的平均值

之间的差异,大的 ES 比小的 ES 更容易检测到,当设计一项实验时,应该选择较高的剂量,但也不应过高,避免出现额外的副作用。检验效能是当 H_0 为假时,实验拒绝它的概率($1-\beta$),通常规定的检验效能为 80%或 90%。显著性水平是实验产生假阳性结果(I 型错误)的概率。在一般实验中通常选择使用双侧检验,但如果反应只可能在一个方向上进行,或者只对某个方向进行研究,那么就可以使用单侧检测。单侧检测可以提供更有说服力的实验,并且需要相对较少的样本量。KISS 确定法中各变量之间关系及部分影响因素见图 1。

4 问题与展望

医学实验研究开展之前,研究人员需要进行广泛的临床前研究来确定药物初步的疗效、毒性与药代动力学。临床前动物研究应密切遵循人类临床试验的既定做法^[16],虽然样本量大小的确定在人类临床研究中得到了充分的强调,但根据已发表的文献来看,样本量大小在动物研究中还并未得到充分的重视^[17]。

表 2 标准效应量值与样本量对应表
Table 2 Correspondence table of SESs value and sample size

样本量 Sample size	80%,单侧 80%,One-tailed text	80%,双侧 80%,Two-tailed texts	90%,单侧 90%,One-tailed text	90%,双侧 90%,Two-tailed texts
4	2.00	2.35	2.38	2.77
5	1.72	2.03	2.02	2.35
6	1.54	1.82	1.80	2.08
7	1.41	1.66	1.63	1.89
8	1.31	1.54	1.51	1.74
9	1.23	1.44	1.41	1.63
10	1.16	1.36	1.32	1.53
11	1.10	1.29	1.26	1.45
12	1.05	1.23	1.20	1.39
13	1.00	1.18	1.15	1.33
14	0.97	1.14	1.10	1.27
15	0.93	1.10	1.06	1.23
16	0.90	1.06	1.02	1.18
17	0.87	1.03	0.99	1.15
18	0.85	1.00	0.96	1.11
19	0.82	0.97	0.93	1.08
20	0.80	0.94	0.91	1.05
21	0.78	0.92	0.89	1.03
22	0.76	0.90	0.86	1.00

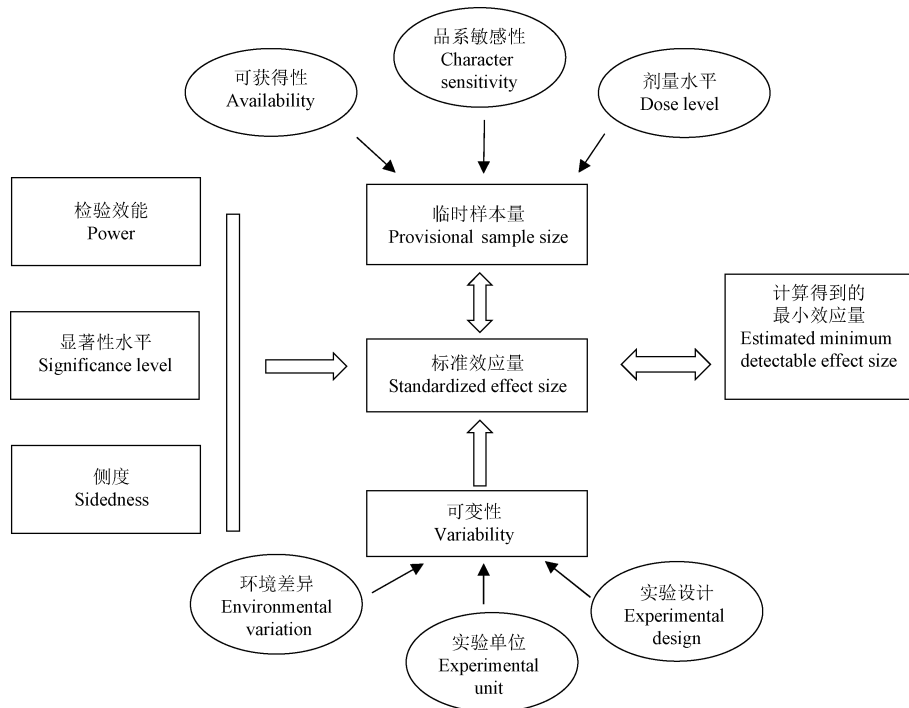


图 1 KISS 确定法中各变量之间关系及部分影响因素

Figure 1 Relationship and influencing factors of variables in KISS approach

动物实验样本量计算的部分假设与条件应在实验开始前确定,但是,由于大多数实验都属于探索性研究,这一情况使得试验的部分条件难以确定,从而导致样本量计算困难的窘境^[18]。实验规划阶段明确样本量不单单对于阐述研究预期目标有重要意义,也可以极大地减少对于动物的伤害以及试验耗材的浪费,实现动物试验流程、步骤优化的同时又遵从了动物伦理准则^[19]。作为动物试验的设计与执行者,需要注意的是没有任何一种样本量计算的方法是万能公式,正如几乎不存在完美的动物模型一样^[20]。有时,确定样本量计算方法有赖于研究人员本身的实践经验以及试验本身的方法与目的。如果一位经验丰富的研究人员对于研究领域非常熟悉,并且已经进行了许多与目标实验类似的实验,那么基于文献与经验的样本量确定方法便可以作为首选。精度分析确定法和两样本均数确定法相对于基于文献与经验的样本量确定方法,因为增加了部分因素和计算,所以,在样本量计算方面具有一定程度的正确性,不再完全依靠经验。资源方程式确定法对于探索全新领域的人来说,提供了一种直接、有效的样本量计算方式,进而避免实验样本的过大或过小。与资源方程式确定法相比,KISS 确定法具有完善的数学证明,但是,它涉及到部分主观因素,必须依靠实验人员来确定最小 ES,

此外,较多变量的存在也会影响精度的准确性。

目前,实验动物样本量的确定尚存在部分问题。专门针对实验动物的统计学样本量确定方法较少,不同于临床患者,实验动物多为近交系,其生理、生化、组织学、形态学上的特征、对药物的反应,甚至行为的类型都趋于一致,因此实验结果一致性好,得出稳健的结论的样本量应该比临床患者所需样本量小。然而,目前基于统计的样本量估计方法多是基于临床研究的,因此需要进行基于实验动物的样本量估算。其次,基于统计学的样本量估算多是针对 1 个效应指标 2 个比较组情况,当实验组数或效应指标增加时,如何选择估算方法报道较少。此外,未获得临床研究样本量的估算同样值得重视,每一项临床试验开展前,都需要经过严格的伦理审查,其中,样本量更是审查的重点之一,而动物实验研究中,多存在对样本量估算不重视或审查不严格的情况。

在未来的医学实验研究中,首先,应该提高研究人员对于医学动物伦理学的了解,培养和传播注重动物福利待遇的观点和理念。其次,研究人员应当加大对统计学、概率学和数学等有利于动物样本量确定内容的学习,拓展样本量计算公式,优化计算方案。此外,成立实验动物样本量计算协会、制定动物样本量确定指南和出版动物样本量计算专

著均可助力医学实验的发展,应当引起科研人员的注意和重视。

简而言之,样本量的选择在医学动物实验中起着举足轻重的作用,但是研究者由于概率论和统计方面知识的缺乏,在动物实验样本大小方面无法进行深入研究,导致样本量的选择具有较大的随机性。随着医学统计学结合实验动物学的发展和传播,动物实验样本量确定方面的研究必将愈加完善,进而为研究过程的严谨性和研究结论的可靠性提供坚定的理论基础和实验依据。

参考文献:

- [1] Percie du Sert N, Hurst V, Ahluwalia A, et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research [J]. *PLoS Biol*, 2020, 18(7): e3000410.
- [2] Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research [J]. *Nature*, 2012, 483(7391): 531-533.
- [3] Basson AR, Cominelli F, Rodriguez-Palacios A. “statistical irreproducibility” does not improve with larger sample size: how to quantify and address disease data multimodality in human and animal research [J]. *J Pers Med*, 2021, 11(3): 234.
- [4] 孙德明, 李蔚鸥, 王天奇, 等. 实验动物福利伦理审查的标准化与我国新国标解读 [J]. *中国比较医学杂志*, 2018, 28(10): 133-137.
- [5] Chitnis KR, Shah AC, Jalgaonkar SV. A study to assess the quality of reporting of animal research studies published in PubMed indexed journals: a retrospective, cross-sectional content analysis [J]. *Cureus*, 2022, 14(1): e21439.
- [6] Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? [J]. *Nat Rev Drug Discov*, 2011, 10(9): 712.
- [7] Lazić SE. Four simple ways to increase power without increasing the sample size [J]. *Lab Anim*, 2018, 52(6): 621-629.
- [8] Kramer M, Font E. Reducing sample size in experiments with animals: historical controls and related strategies [J]. *Biol Rev*, 2017, 92(1): 431-445.
- [9] Charan J, Kantharia ND. How to calculate sample size in animal studies? [J]. *J Pharmacol Pharmacother*, 2013, 4(4): 303-306.
- [10] ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International conference on harmonisation E9 expert working group [J]. *Stat Med*, 1999, 18(15): 1905-1942.
- [11] Ko MJ, Lim CY. General considerations for sample size estimation in animal study [J]. *Korean J Anesthesiol*, 2021, 74(1): 23-29.
- [12] Bacchetti P, Leung JM. Sample size calculations in clinical research [J]. *Anesthesiology*, 2002, 97(4): 1028-1032.
- [13] Novick S, Zhang T. Mean comparisons and power calculations to ensure reproducibility in preclinical drug discovery [J]. *Stat Med*, 2021, 40(6): 1414-1428.
- [14] Arifin WN, Zahiruddin WM. Sample size calculation in animal studies using resource equation approach [J]. *Malays J Med Sci*, 2017, 24(5): 101-105.
- [15] Festing MF. On determining sample size in experiments involving laboratory animals [J]. *Lab Anim*, 2018, 52(4): 341-350.
- [16] Muhlhausler BS, Bloomfield FH, Gillman MW. Whole animal experiments should be more like human randomized controlled trials [J]. *PLoS Biol*, 2013, 11(2): e1001481.
- [17] 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.
- [18] Cesana BM, Antonelli P. Sample size calculations in clinical research should also be based on ethical principles [J]. *Trials*, 2016, 17(1): 149.
- [19] Richter V, Muehe R, Mayer B. How much confidence do we need in animal experiments? Statistical assumptions in sample size estimation [J]. *J Appl Anim Welf Sci*, 2018, 21(4): 325-333.
- [20] 李磊, 刘建勋, 任钧国, 等. 中医药动物模型研究现状及展望 [J]. *中国比较医学杂志*, 2022, 32(1): 104-110.

[收稿日期] 2022-06-09