



孙 强, 研究员, 博士生导师。中国科学院脑科学与智能技术卓越创新中心(神经科学研究所)非人灵长类研究平台主任。国家杰出青年基金获得者, 科技部中青年科技创新领军人才, 荣获2018年谈家桢生命科学创新奖、2019年国务院政府特殊津贴、2020年药明康德生命化学研究奖-学者奖、2022年度上海市自然科学奖一等奖。中国实验动物学会理事, 中国实验动物学会灵长类实验动物专业委员会常务理事, 上海市实验动物学会理事, 上海市实验动物学会生物安全专业委员会委员。长期致力于实验动物管理和模式动物构建及相关技术的研发工作, 为了向公众普及实验动物相关知识, 坚持实验动物科普写作多年, 为“实验动物那些事儿”公众号主要作者之一, 发表科普文章70余篇。

如果动物实验不严谨

孙 强

(中国科学院脑科学与智能技术卓越创新中心, 上海 200031)

[中图分类号] R-332; Q95-33 [文献标志码] B [文章编号] 1674-5817(2024)02-0234-04



本栏目的第一篇文章中, 通过几个较早的动物实验案例, 以“为什么要开展动物实验”为题, 向读者朋友们介绍了动物实验对人类和动物健康的贡献^[1]。如今, 虽然通过使用动物开展实验来评价候选药物的疗效、代谢及潜在毒性是新药研发不可或缺的环节, 但科学界也会有不同的声音出现。尤其是基于动物实验结果的候选药物临床转化率仅有11%时^[2], 有些人认为动物实验完全没有必要。事实果真如此吗? 如果药物进入临床前没有通过动物实验对候选药物进行安全性评价, 不知将会有多少患者受到伤害。在新药研发过程中, 不仅要开展动物实验, 而且要进行科学严谨的动物实验。如果动物实验设计和实施过程不严谨, 同样会出问题。例如因出现严重不良反应而被召回的“新药”反应停和万络, 之所以在上市前没有发现其毒性和不良反应, 不严谨的动物实验结果难辞其咎。下面就以几个经典案例来介绍不严谨的动物实验所带来的危害。

1 反应停事件

约60年前, 一款缓解孕妇晨吐的镇静药沙利度胺(Thalidomide, 商品名为反应停)在上市5年后, 因被发现可导致婴儿海豹胎(四肢短得就像海豹的鳍足)和其他畸形而被叫停使用。反应停是20世纪50年代由联邦德国格兰泰公司投入开发的一种镇静类药物, 其声称该药能够显著抑制孕妇晨吐等妊娠反应。1957年10月, 反应停正式投放欧洲市场, 不久后进入日本市场, 并在此后不到一年的时间内风靡欧洲、日本、非洲、澳大利亚和拉丁美洲。但在美国申请上市许可时, 遇到了美国食品药品监督管理局(Food and Drug

Administration, FDA)仔细且繁琐的市场准入调查, 最终没能得到进入美国市场的许可, 并促成“美国国会于1962年通过法案, 加强对药物的管理, 授予FDA更多的权力, 要求新药在获准上市前必须经过严格的试验, 并提供药物不良反应和中长期毒性的数据, 必须对至少两种妊娠动物进行致畸性试验”^[3]。

到1961年, 澳大利亚产科医生威廉·麦克布里德(William McBride)向英国《柳叶刀》(*The Lancet*)杂志投稿, 以“Letter to Editor”的形式报道了反应停能导致婴儿畸形的现象^[4]。在麦克布里德接生的产妇中, 许多人产下的婴儿患有罕见的海豹胎畸形症状, 而这些产妇都曾经服用过反应停。随后, 格兰泰公司撤回联邦德国市场上所有反应停, 不久其他国家也相继停售, 但此时已经有万余名因服用反应停而导致畸形的胎儿出生。由于这一事件, 格兰泰公司支付了1.1亿西德马克的赔偿, 并破产倒闭。

回顾发现, 在反应停上市前, 格兰泰公司使用大鼠开展了动物实验研究, 但并没有发现畸形大鼠出生。后续调查证明, 他们的动物实验设计非常不严谨。首先, 沙利度胺在大鼠体内的半衰期约为0.5 h, 而在人体内的半衰期则有7.3 h之久。因此, 沙利度胺在大鼠体内会很快被代谢掉。其次, 抗氧化剂如谷胱甘肽可以抑制沙利度胺引起的细胞凋亡, 而大鼠比人类具有更多的抗氧化剂, 这些抗氧化剂可保护它们的胚胎免受沙利度胺引起的破坏^[5]。因此, 可以说大鼠并不是一种可以充分评估沙利度胺毒性的实验动物, 相较而言, 使用非人灵长类动物则很容易观察到沙利度胺对胚胎发育的影响。

另外一个容易被忽略的因素是, 大鼠为多胎生产

实验动物，其分娩又多发生在夜间，且母鼠有吃掉发育畸形仔鼠的习惯。因此，当实验人员白天检查大鼠生产情况时往往只能观察到健康的出生个体。虽然窝平均产仔数可能有少许减少，但通常也不会引起注意。后续通过剖宫产和夜间观察大鼠产仔，确实发现在妊娠特定阶段给予沙利度胺，大鼠也可以产下畸形的胎儿。也就是说，如果当时格兰泰公司在开展动物实验时能够考虑到这个细节，改为剖宫产或想办法直接观察大鼠产仔时的情况，或许就可以观察到反应停的不良后果，从而避免悲剧的发生。而且格兰泰公司于1956年发表的沙利度胺动物实验相关文章中显示其动物实验既没有设置安慰剂组，也没有交代给药持续时间，更没有进行双盲对照实验设计^[6]。如此不严谨的实验设计，其团队注意不到动物实验的细节问题也就一点都不意外了。综上，正是由于沙利度胺在上市前的动物安全性评价实验存在严重缺陷，人们没能及时发现其不良反应，最终导致了反应停悲剧的发生。

2 万络事件

如果说反应停事件还只是新药研发和审批不健全时代的产物，那么万络的匆忙上市和退市则是在利益驱动下，当代药物研发中由于动物实验不严谨而造成严重后果的经典案例。

身体在经历发烧、疼痛或炎症反应时，往往会通过抑制或阻断身体内不同部位的前列腺素分泌来缓解症状。例如，退烧药布洛芬和对乙酰氨基酚就是通过抑制下丘脑前列腺素合成来调控体温调节中枢内的体温调定点，促使其恢复正常而实现退烧的；阿司匹林则通过抑制外周前列腺素合成来实现其镇痛、退热和抗炎的功效。上述这些具有抑制或阻断身体不同部位前列腺素分泌或功能的药物属于非选择性环氧合酶（cyclooxygenase, COX）（又称前列腺素内氧化酶还原酶）抑制剂。它们都是通过非特异性地抑制或阻断COX活性，实现其抑制前列腺素合成的功能，进而发挥药效。

COX是催化花生四烯酸转化为前列腺素的关键酶，有两种同工酶：COX-1和COX-2。COX-1除了下丘脑外，还在血管、胃、肾等多种组织中表达，参与血管舒缩、血小板聚集、胃黏膜血流、胃黏液分泌及肾功能等的调节，具有保护胃肠黏膜、调节血小板聚集、调节外周血管阻力和调节肾血流量分布的功能。因此，这些非选择性COX抑制剂类药物在发挥其药效的同时，也会抑制广泛表达在身体其他部位的COX-1，从

而产生不良反应。最典型的不良反应就是引起胃部不适，甚至出现胃溃疡和胃出血等。COX-2则是诱导表达型，只有在身体经历了诸如化学和机械损伤或微生物感染后才会激活。因此，如果能开发出一种药物可以选择性地阻断COX-2的功能，就有望避开上述非选择性COX抑制剂类药物的不良反应。

20世纪90年代，制药界的巨头之一德国默克公司启动了寻找COX-2特异性抑制剂的研究。他们很快推出了一款新药罗非昔布，商品名为万络。并通过8 000多名患者的临床试验证明，与当时广泛使用的同类药物萘普生相比，该药在治疗类风湿性关节炎时具有更少的胃肠道不良反应，差不多将胃肠道事件的风险降低了一半。但由于过于关注胃肠道不良反应，默克公司在研发万络的过程中，有意或无意地忽略了该药对心血管系统的严重不良反应，且在不断出现心血管不良反应报告的情况下仍没有重视。直到默克公司继续研究万络在预防腺瘤性息肉等方面的其他适应证时，证实其存在引发心血管的风险，该药才从市场上撤回。但此时仅美国就约有400万人服用了万络。美国FDA的一名调查员估计，万络可能导致约14万人罹患心脏病，其中6万人或将因此死亡^[7]。

排除其他各种涉及公司经济利益等原因外，不严谨的动物实验结果也是万络能够最终通过审查上市的原因之一。早在万络上市之前，宾夕法尼亚大学医学博士加勒特·菲茨杰拉德（Garret FitzGerald）领导的研究小组就观察到万络抑制了前列环素生成。鉴于前列环素对心脏具有潜在的保护作用，加勒特推测该药可能有导致心脏病和中风的风险。但这并没有引起默克公司的足够重视，他们也没有就此开展更为严谨的动物实验来分析万络在心血管方面的风险。而在该药退市后，更进一步的研究结果显示，选择性地敲除血管内皮细胞中的COX-2会使小鼠出现血栓和高血压，这些小鼠就像服用万络的人类一样，前列环素也出现了下降^[8]。想象一下，如果默克公司在万络上市前就能使用COX-2敲除小鼠开展一个并不需要花费太长时间和精力的研究，或许就可以挽救成千上万人的生命。

3 渐冻人症药物研发

因中枢神经系统内控制骨骼肌运动的神经元退化导致的肌萎缩性脊髓侧索硬化症（amyotrophic lateral sclerosis, ALS，也称渐冻人症）是一种渐进且致命的神经退行性疾病。由于运动神经元退化和死亡，肌肉逐渐萎缩致大脑完全丧失控制随意运动的能力，最终

会造成发音、吞咽、呼吸和运动障碍。著名物理学家斯蒂芬·霍金 (Stephen Hawking) 就是一名 ALS 患者。目前还没有真正可有效治疗 ALS 的药物和方法。

绝大多数 (90%~95%) 的 ALS 发病原因不明,但也有约 5%~10% 的 ALS 患者表现出了家族遗传性。其中,超氧化物歧化酶 1 (superoxide dismutase 1, *SOD1*) 基因突变约占这些遗传性 ALS 病例的 10%。*SOD1* 也是第一个被鉴定出来的 ALS 风险基因^[9]。因此很快就有科学家构建了一种含有第 93 位甘氨酸突变为丙氨酸 (G93A) 的人 *SOD1* 基因的转基因小鼠模型,全名为 B6SJL-Tg (*SOD1**G93A) 1Gur/J 转基因小鼠^[10]。该转基因小鼠的基因组中被转入了 23 个拷贝的人源 *SOD1**G93A 突变基因,接下来我们将其简称为 *SOD1**G93A 转基因小鼠。

自 1996 年 *SOD1**G93A 转基因小鼠问世以来,其就被广泛应用于 ALS 药物研发的临床前动物实验中。到 2004 年,短短 8 年时间内,已至少有 50 篇文章描述了从小分子到病毒载体的各类 ALS 治疗方案可延长 *SOD1**G93A 转基因小鼠的寿命,其中最多可以延长 100 d (约占原寿命的 78%)。因此,这些药物很快进入临床试验。然而,迄今为止,除了利鲁唑最多可延长 ALS 患者两个月寿命之外,没有任何一种药物显示出对人类 ALS 疾病的进程有显著影响^[11]。这给医药公司带来了巨大损失。例如,2008 年在《美国国家科学院院刊》(*PNAS*) 上发表的一篇文章显示,锂可以将 *SOD1**G93A 转基因小鼠的存活时间延长 30 d^[12];由于锂早已经被用于治疗精神分裂症,许多 ALS 患者开始超说明书服用该药物,希望减缓 ALS 的疾病进程;然而,招募了数百名患者,花费了远超 1 亿美元的 3 个同时启动的独立 III 期临床试验结果却显示,锂对 ALS 无任何治疗效果。

其实不仅仅是 ALS 的药物研发,在整个医药研发领域,动物实验结果的临床转化率也是非常低的,平均在 11% 左右^[1,13]。那么到底是什么原因导致了如此多的动物实验结果无法实现在人类的临床转化呢?

1999 年,美国麻省理工学院的机械工程师詹姆斯·海伍德 (James Heywood) 在得知他的兄弟斯蒂芬·海伍德 (Stephen Heywood) 被确诊患有 ALS,并无任何有效治疗方法后,与家人一起创立了一家非营利性生物技术公司,即 ALS 治疗开发研究所 (ALS-TDI)。ALS-TDI 成立后,开展了一项旨在寻找 *SOD1**G93A 转基因小鼠实验结果临床转化失败原因的项目,即使用 *SOD1**G93A 转基因小鼠重复已发表科研

论文中的动物实验,以找出其临床转化失败的原因。

在 ALS-TDI 开展的重复性实验中,他们不仅严格执行盲法对照,而且对给药组和对照组的动物进行了更为精细地分配和动物筛查以排除系统误差。他们的分组及实验要求包括:(1) 年龄匹配;(2) 性别匹配;(3) 同窝匹配,即使用同一天出生的同一非转基因母亲和转基因父亲的后代进行分组,具体来说,治疗组中的每个雄性 (和雌性) 动物在对照组中都有一个同窝兄弟 (和姐妹);(4) 样本量设定,每组使用动物数量为 24 只 (12 雌, 12 雄);(5) 死因筛查,检查记录每一只动物的死亡原因,排除非 ALS 致死个体;(6) 转基因拷贝数监控,有研究证明 *SOD1**G93A 转基因小鼠在传代过程中会出现转基因拷贝数丢失的现象,且低拷贝的转基因小鼠的平均寿命会延长。

在这一严格的实验分组和动物筛查条件下,ALS-TDI 的科研人员总共使用了 5 429 只 *SOD1**G93A 转基因小鼠测试了包括利鲁唑在内的 100 多种候选化合物 (大部分已有论文发表并证明可延长 *SOD1**G93A 转基因小鼠寿命),结果却发现没有一种化合物能表现出可延长 *SOD1**G93A 转基因小鼠寿命的效果^[11]。

为了找出那些已经发表的实验结果无法被复现的原因,研究人员以在 ALS-TDI 开展实验中所使用的 2 000 多只对照小鼠作为研究对象,进行了计算机分组模拟实验。由于这些小鼠从出生到死亡的信息以及死亡原因在 ALS-TDI 都有详细的记录,因此他们能够以所有对照小鼠为数据库,分析前述重复实验中设定的不同分组和动物筛查条件对实验结果的影响。即通过计算机对这些没有经过任何药物处理的小鼠根据前述重复实验中设定的不同分组和动物筛查条件进行随机分组,将一组设为假定的实验组,而另一组设为假定的对照组,然后依据已有的小鼠死亡时间记录,统计两组小鼠之间死亡时间是否存在显著差异。经过多次重复,计算出现 5% 以上显著差异 (统计学上通常设置的显著置信区间,即 $P < 0.05$) 的百分比。由于这些小鼠没有接受过任何化合物处理,因此如果假定的实验组和对照组小鼠死亡时间出现了显著性差异,那么其原因应该是分组时的系统误差。

结果显示,在样本量大小为 4 的计算机分组模拟对比实验中,随机分组和不进行死因筛查 (排除和不排除低转基因拷贝数个体) 条件下,两组间出现显著差异的可能性超过了 50%。而在随机分组的同时进行死因筛查并排除低转基因拷贝数个体和进行性别同窝匹配的条件下,样本量为 4 的模拟分组实验中也有

30% 的可能出现两组死亡时间有显著差异的情况。随着样本量的增加, 出现显著性差异的可能性逐渐减少, 但要实现低于 5% (即 $P > 0.05$) 的非显著性差异, 样本量要达到每组 24 只动物。如果不做分组条件控制, 即使样本量达到 50 只也难以排除系统误差^[11]。

4 启示

反应停事件和万络事件告诉我们: 药物研发期间的动物实验不仅不可或缺, 而且还需要选择最合适的实验动物 (如反应停应使用非人灵长类动物, 万络应补充使用基因敲除小鼠); 同时应对任何潜在的风险因素都给予重视, 并通过严谨和充分的动物实验加以检测和验证。

如果说这两个案例主要反映了制药公司在药物研发过程中, 临床前动物实验不严谨会给患者带来严重危害, 那么渐冻人症药物研发案例则提示科学研究中动物实验不严谨所带来的更多危害。

ALS-TDI 的计算机模拟实验结果说明, 前述那些已经发表的基于 SOD1*G93A 转基因小鼠的实验结果无法实现临床转化的真正原因极大可能是源于其实验设计的不合理, 即这些实验因为没有排除分组时的系统误差而导致了假阳性结果。这个研究也暴露了当下使用实验动物开展研究的诸多问题, 诸如: 盲目地追求减少使用动物的数量; 对所用动物的遗传、微生物和繁殖等特性不了解; 实验过程中对动物关注不够, 不清楚动物发病或死亡的真正原因等。这些问题不仅需要引起实验动物从业人员的注意, 而且需要科研人员同样给予关注。因为实验设计不严谨导致的假阳性结果不仅是人力、物力和财力的巨大浪费, 也严重有违动物实验伦理福利原则, 这样的实验不会带来科学进步, 反而浪费了大量实验动物的生命。比如在 ALS-TDI 开展的重复性实验中, 那些无法重复出阳性结果但已经发表的研究工作累计使用小鼠的数量已超过了 18 000 只。

此外, 在动物实验伦理申请时常用替代、减少和优化原则 (俗称 3R 原则) 来指导伦理审查工作, 有时为了满足 3R 原则中的“减少”原则, 加上开展实验的成本考虑, 往往选择使用最少的动物数量 (比如多为每组 6 只)。但从 ALS-TDI 的计算机分组模拟对比研究看, 样本量过小会导致系统误差加大, 因此极可能出现假阳性结果。这个时候大家或许忘了 3R 原则使用的大前提是要能够实现拟开展实验的科学目标, 如果单纯为了“减少”而使用更少的实验动物导致科学目标

无法实现, 那就不是真正的减少, 而是对动物生命的漠视 (实际的浪费)。

综上, 3 个案例告诉我们, 无论是科学研究, 还是候选药物临床前的安全性评价, 都需要开展严谨的动物实验, 否则不仅会令所得科研成果无效而浪费动物的生命, 也会造成巨大的经济损失, 并给患者带来严重的健康风险。

(2024 年 3 月 4 日星期一, 于强业路 500 号)

【参考文献 References】

- [1] 孙强. 为什么要开展动物实验[J]. 实验动物与比较医学, 2024, 44(1): 121-126. DOI: 10.12300/j.issn.1674-5817.2023.177.
- [2] HAY M, THOMAS D W, CRAIGHEAD J L, et al. Clinical development success rates for investigational drugs[J]. Nat Biotechnol, 2014, 32(1):40-51. DOI: 10.1038/nbt.2786.
- [3] 方舟子. “反应停”悲剧[EB/OL]. (2005-04-06)[2024-03-03]. http://zqb.cycl.com/content/2005-04/06/content_1063523.htm.
- [4] YANG T J, YANG T S, LIANG H M. Thalidomide and congenital abnormalities[J]. Lancet, 1963, 1(7280): 552-553. DOI: 10.1016/s0140-6736(63)91347-3.
- [5] VARGESSON N. Thalidomide-induced teratogenesis: history and mechanisms[J]. Birth Defects Res C Embryo Today, 2015, 105(2):140-156. DOI: 10.1002/bdrc.21096.
- [6] BOTTING JH. Animals and medicine: The contribution of animal experiments to the control of disease[EB/OL]. [2024-03-03]. <https://books.openbookpublishers.com/10.11647/obp.0055.pdf>.
- [7] KRUMHOLZ H M, ROSS J S, PRESLER A H, et al. What have we learnt from Vioxx? [J]. BMJ, 2007, 334(7585): 120-123. DOI: 10.1136/bmj.39024.487720.68.
- [8] YU Z, CRICHTON I, TANG S Y, et al. Disruption of the 5-lipoxygenase pathway attenuates atherogenesis consequent to COX-2 deletion in mice[J]. Proc Natl Acad Sci U S A, 2012, 109(17):6727-6732. DOI: 10.1073/pnas.1115313109.
- [9] ROSEN D. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis[J]. Nature, 1993, 364(6435):362. DOI: 10.1038/362059a0.
- [10] TU P H, RAJU P, ROBINSON K A, et al. Transgenic mice carrying a human mutant superoxide dismutase transgene develop neuronal cytoskeletal pathology resembling human amyotrophic lateral sclerosis lesions[J]. Proc Natl Acad Sci U S A, 1996, 93(7):3155-3160. DOI: 10.1073/pnas.93.7.3155.
- [11] 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.
- [12] FORNAI F, LONGONE P, CAFARO L, et al. Lithium delays progression of amyotrophic lateral sclerosis[J]. Proc Natl Acad Sci U S A, 2008, 105(6): 2052-2057. DOI: 10.1073/pnas.0708022105.
- [13] KOLA I, LANDIS J. Can the pharmaceutical industry reduce attrition rates? [J]. Nat Rev Drug Discov, 2004, 3(8): 711-715. DOI: 10.1038/nrd1470.

(收稿日期: 2024-03-07 修回日期: 2024-03-24)

(本文编辑: 张俊彦, 富群华)

【引用本文】

孙强. 如果动物实验不严谨[J]. 实验动物与比较医学, 2024, 44(2): 234-237. DOI: 10.12300/j.issn.1674-5817.2024.039.