

如何正确把握实验设计的三要素

胡良平,王琪,毛玮,陶丽新

(军事医学科学院生物医学统计学咨询中心,北京 100850)

Zhongguo Gushang/China J Orthop & Trauma, 2009, 22(4): 324-326 www.zggszz.com

任何一项实验研究通常都包括受试对象、实验因素和实验效应三个部分,故称其为实验设计三要素。医学实验的受试对象主要是人、动物和样品。进行实验研究,一个很明显的特点就是要向受试对象施加干预措施,在统计学上称其为因素(如药物种类、手术方式或其他治疗手段等),通过对某些指标在不同受试对象身上的表现,得到所谓的实验结果,通过运用某些合适的统计学方法找到因素不同水平或不同因素作用条件下观测指标取值的变化规律性,从而达到了解和掌握人们关心的事物或现象的本质,以实现研究者最初的研究目的。

例 1:《锚定法单开门椎管成形术对术后轴性症状和颈椎曲度影响的病例对照研究》^[1]一文,选择锚定法单开门椎管成形术并获得完整随访的患者 56 例(锚定单开门,A 组),与 50 例同期进行单开门手术的患者(传统单开门,B 组)进行对比分析,记录 JOA(17 分法)评分等。资料见表 1。

表 1 两组手术前后 JOA 评分变化 ($\bar{x} \pm s$)

Tab.1 The comparison of changes of JOA scores before and after operation between two groups ($\bar{x} \pm s$)

组别	例数	术前(分)	术后(分)	改善率(%)
A 组	56	6.3±3.0	12.9±2.7	61.1±24.5
B 组	50	6.1±2.9	11.9±2.4	53.3±23.3

从本例可以看出,实验所用的两种手术方法为实验因素,该病患者为受试对象,JOA 评分为观测指标,体现实验因素所产生的实验效应。

1 受试对象

根据实验目的确定实验因素之后,接着就是选择什么样的受试对象。受试对象是指承受实验因素作用的对象。

1.1 受试对象的选取 根据实验研究的目的和具体情况,应结合专业知识选择合适的受试对象。从统计学角度看,更多场合下是应当结合具体问题制定出“纳入和剔除”受试对象的标准,以减少或消除重要非实验因素对实验结果的干扰和影响。受试对象的同质性必须予以高度重视。在药物的临床实验中,还必须考虑到“伦理道德”、“尊重人权”和“提高受试者依存性”等问题^[2]。

1.2 受试对象的种类 根据受试对象的特征和属性可将其粗分为生物体和非生物体两大类,生物体又可粗分为人体和非人体两类,其中人体又可细分为患者和非患者,非人体又可细分为动物、植物,另外,还可以是人体的离体标本或细胞等。

不同的实验目的、实验因素、实验的技术水平需要选取不同的受试对象。**①动物。**例 2:《悬吊模拟失重及解悬吊对大鼠骨密度及生物力学的影响》^[3],20 只 Wistar 雄性大鼠随机分成空白组和模型组。模型组尾部悬吊 14 d,解悬吊后继续饲养 14 d,空白组则正常饲养 28 d。实验第 28 天处死大鼠,检测右侧股骨及 L₄ 椎体 BMD 及生物力学强度,差异有统计学意义。资料见表 2。**②某病患者。**例 3:《旋转扳法和穴位推拿治疗青年颈性眩晕的随机对照实验》^[4],76 例表现为椎基底动脉异常高流速状态的青年颈性眩晕患者随机分为旋转扳法组(A 组,38 例)和穴位推拿组(B 组,38 例),分别施加相应的治疗手法。应用经颅多普勒超声(TCD)检测手法前后椎基底动脉流速并观察流速改变和临床疗效。资料见表 3。**③离体标本。**例 4:《前下后上型齿突骨折螺钉内固定的生物力学研究》^[5],收集 20 具新鲜成人上颈椎标本,分为 3 组,模拟齿突 II 型骨折模型,骨折线为前上后下,截骨角度分别为 0° 6 具,17° 8 具,25° 6 具,并进行齿突螺钉固定。进行旋入力矩、最大轴向拔出力和刚度测试。各组最大旋入力矩、抗拔出差异无统计学意义。资料见表 4。

1.3 受试对象的数量 受试对象的数量通常指实验研究中

表 2 两组动物实验第 28 天离体股骨、腰椎骨密度的变化 ($\bar{x} \pm s, g/cm^2$)

组别	鼠数	股骨	L ₄ 椎体
空白组	10	0.153 5±0.010 2	0.143 8±0.007 9
模型组	10	0.135 3±0.011 2	0.128 9±0.006 9

表 3 两组治疗后 6 个月疗效比较 (例)

组别	例数	治愈	显效	有效	无效
旋转扳法组	38	13	7	8	10
穴位推拿组	38	20	12	3	3

表 4 不同截骨度与各测量参数值 ($\bar{x} \pm s$)

组别	最大旋入力矩(N·m)	骨折移位 (mm)
0°截骨组	4.53±2.59	0.70±0.61
17°截骨组	6.18±2.22	1.61±0.64
25°截骨组	4.47±1.77	2.88±0.80

总共需要多少样本含量,也称样本大小,在统计学上称为“样本大小估计问题”。值得注意的是,若实验涉及的因素很多,根

据因素的水平组合会形成很多小组,每个小组中的受试对象的数量不能太小。样本大小估计是一个比较复杂的问题,它涉及到实验设计的类型、观测指标的性质、有关的先验知识和对结果精确度的要求。比较好的做法是根据具体情况利用相应的计算公式估算出样本含量^[6]。

1.4 选择受试对象的基本要求 选择受试对象时应着重考虑下列基本要求:①敏感性:受试对象对被施加的处理因素应有较高的敏感性,容易显示效应。②特异性:受试对象对被施加的处理因素应有较强的特异性,排除非处理因素干扰。③稳定性:受试对象对被施加的处理因素的反应有较大的稳定性,减少误差。④同质性:研究对象的种属、生物学特性及其他条件保持均衡。⑤依从性:受试对象接受处理因素的合作程度。⑥经济性:指受试对象容易找到,费用低。

1.5 实际工作者在确定受试对象方面常犯的错误 ①在以动物为实验研究的场合中,常盲目选取某种受试对象。②在以人为受试对象的实验研究中,常忽视伦理道德问题,选择不恰当的对照组。

例如,在制作不同疾病的动物模型时,必须依据专业知识和以往的经验,选择最合适的动物作为受试对象。否则,实验很容易失败。如制作糖尿病动物模型适合选用自发性高血压大鼠;制作高血脂或动脉粥样硬化动物模型适合选用日本大耳白兔。

又例如,做染色体畸变实验与分析时,由射线照射引起染色体畸变,观察指标:“双着色粒,染色体环”的发生率。发生率越高,表明剂量越大,在一定剂量范围内,剂量与效应之间呈线性关系。关键在于选用什么细胞作为受试对象,是选外周血淋巴细胞还是选骨髓细胞?国际上公认用“外周血淋巴细胞”,因为淋巴细胞相对比较一致、均匀,且好采样。后来,有人为了“创新”,提出用“骨髓细胞”来培养。众所周知,骨髓细胞中有红细胞系统、粒细胞系统、血小板系统,有淋巴细胞,还有其他细胞,不仅细胞系统多、杂,而且还同时存在处于幼、早、中、晚期的细胞,因此,这样的细胞体系不能很好地反映剂量与效应之间的线性关系。事实证明,不宜用这样的细胞体系来做此类实验研究的受试对象!

2 实验因素

实验因素就是研究者希望通过实验考察其对实验结果是否有影响的性质不同的不同实验条件的总称,包括物理、化学、生物、社会因素等。研究对象自身具有的、可影响其发展过程的某些因素(如性别、年龄等),也可作为实验因素。同一种因素还可能不同水平,如:药物的浓度高低、时间的长短、温度的高低等。

例 5:《电针治疗膝关节炎临床对照实验》^[7],将 40 例膝关节炎患者按就诊顺序随机分为 2 组,每组 20 例。电针组男 8 例,女 12 例,平均年龄(62.50±9.23)岁,平均病程(9.85±8.17)年,采用电针治疗。对照组男 7 例,女 13 例,平均年龄(61.15±6.75)岁,平均病程(8.60±6.89)年,口服双氯芬酸钠治疗。采用 Lysholm 膝关节评分标准评定治疗前后膝关节功能。其中电针为物理实验因素,双氯芬酸钠为化学实验因素。

例 6:《微创路径椎间盘髓核摘除术治疗腰椎间盘突出症》^[8],2000 年 11 月至 2006 年 4 月对 61 例腰椎间盘突出症

患者利用切吸设备中的穿刺针和套管等器械,进行微创路径椎间盘髓核摘除术治疗,对 61 例手术病例术后 0.5、1、6 个月及 1 年进行随访,观察疗效。资料见表 5。

可见,实验因素为“术后时间”,根据时间的长短分为 4 个

表 5 术后随访时间、例数及治疗结果

术后时间 (月)	疗效(例)				优良率 (%)
	优	良	可	差	
0.5	38	14	7	2	85.2
1	44	13	3	1	93.4
6	45	12	3	1	93.4
12	48	10	2	1	95.0

水平,分别为术后 0.5、1、6 个月及 1 年。

实验因素依据研究目的而确立,其选择是否恰当,直接影响研究过程和最终结果,因此,实验因素的实质内容必须清楚。影响实验结果的因素很多,实验设计时应注意。

2.1 要抓住实验中的主要因素 任何实验效应都是多种因素综合作用的结果,我们不可能也没有必要将所考虑到的一切因素都放在一次或几次实验之中进行观察和分析,只能抓住对实验结果影响较大的主要因素。例如研究某种骨病的治疗方法,与其有关的因素有很多,如年龄、病程、损伤程度等,其中每个因素又分若干水平。如果我们选定其中的两个因素各有 10 个水平,就有 $10^2=100$ 种实验条件;若选定 4 个因素,各取 10 个水平,那就有 10 000 种实验条件。每种实验条件下至少要做 2 次独立重复实验(具体应重复几次,应有根据地估算出来),因此,一个好的实验设计,首先要求在众多因素与水平中抓住主要的几个^[9],水平数也不能定得太多。

2.2 注意区别实验因素与非实验因素 在实验研究中,除了实验因素对观测结果有影响外,还会有其他因素对观测结果产生影响,如受试对象的性别、体重、年龄、心理状态等基本特征,实验者的技术水平和情绪等是否稳定,药物或试剂的质量和性能是否稳定,做实验时的环境和条件是否稳定等,它们都被称为非实验因素。非实验因素虽然不是实验因素,但由于其中有些会影响实验结果,产生混杂效应。所以非实验因素又称为混杂因素,因为它会影响实验结果,又称干扰因素。设计时应明确这些非实验因素,有意识地控制和消除其干扰作用。应结合具体问题,选定其中对观测结果具有最大影响的一个或若干个,将其定为“重要的非实验因素(简称为区组因素)”,在实验设计时,设法对它们加以控制,这就是所谓“含区组因素”的实验设计类型,如配对设计、随机区组设计、平衡不完全随机区组设计、含区组因素的析因设计等。更多的时候是将受试对象按重要实验因素进行分组,然后再将每组中的受试对象完全随机地均分入实验组与对照组中去,此为“分层随机化”。若重要的非实验因素是定量的,则应从实验开始之时,就记录每个受试对象的定量影响因素的数值大小,将来进行数据处理时采用类似“协方差分析”等技术将其影响降至最低程度。

例 7:《关节外侧小切口治疗肱骨髁上骨折》^[10],肱骨髁上骨折 58 例,男 42 例,女 16 例,年龄 3~65 岁,平均 9.4 岁。术后 58 例随访 1~5 年,平均 2.4 年,依据 Flynn 临床功能评定标准:优 48 例,良 6 例,可 3 例,差 1 例。

释疑;根据常识,患者的年龄、性别等对术后疗效具有重要影响。应在实验设计时,结合专业知识,考虑重要的非实验因素对实验结果可能带来的影响,以免因考虑不周,到实验结束时无法弥补而导致整个实验的失败。本资料在实验设计时,应考虑不同年龄、性别、随访时间长短对术后疗效可能带来的影响,可通过采用合适的多因素实验设计类型及其相应设计下定量资料的协方差分析等技术方法,消除这些重要非实验因素所造成的不利影响。

2.3 注意实验因素间的交互作用 多种因素之间往往存在交互作用。交互作用是在原效应的基础上产生增强效应或减弱效应的作用,最易于理解的是两药共用可发生药效的协同作用或拮抗作用。事实上,不仅两种药物同时使用可能会产生交互作用,非药物的实验因素之间也常可产生交互作用。所谓交互作用,就是一个因素不同水平对观测结果的影响情况会随着另一个因素水平的改变而改变。因此在实验设计时除了要考虑各因素单独施加于受试对象的效应,还应特别关注因素配合施加于受试对象后产生的效应。

2.4 实验因素要标准化 实验因素标准化就是如何保证实验因素在整个实验的过程中始终如一,保持不变。例如实验因素是药品种类,除应确定药品的名称、性质、成分、作用及用法外,还应明确生产厂家、药品批号、出厂日期及保存方法等。实验因素水平的定义与实际不符,所得结果当然就不同了,此时很容易得出错误的结论。同样,其他非实验因素也必须遵循标准化原则(即设法使它们在实验组与对照组间达到高度均衡),否则这些非实验因素或多或少会影响研究结果,导致结论出现较严重的偏差。因此,在实验设计时应使实验因素标准化,并制订出标准化的具体措施和方法。

2.5 实际工作者在确定实验因素方面常犯的错误 ①喜欢用“组别”和“处理”这些词作为全部实验分组的总称,无论实验中涉及到多少个组,他们总把“组别”或“处理”理解为一个实验因素,很自然地将其下的各组理解为该因素的各个水平。②毫无根据的选择很多实验因素,使实验研究变得十分复杂,往往因为实验条件、人力和时间都达不到起码的要求,使实验研究半途而废或因安排不当,组间夹杂着某些混杂因素的影响,导致结论可信度低。③由于研究者实验设计水平贫乏,不会科学地利用多因素实验设计技术,只能将其他实验因素视而不见,每次只让自己关心的某个实验因素取不同的水平,当实验因素之间不独立时,常得出错误结论。

3 实验效应

实验效应就是实验因素作用于受试对象后所产生的结果。它是通过具体的观测指标来实现的,即观测指标是用来反映实验因素作用强弱的重要“尺子”,必须结合实验因素的性质和特点、仪器、试剂和技术水平等多方面综合考虑,找出“特异性强、灵敏度高、准确、可靠”的观测指标,以“客观指标”为主,以“半客观和主观指标”为辅。

3.1 指标的选择 指标反映实验效应,这些指标包括定性指标和定量指标等。指标的选择必须注意它的针对性,即选用的指标必须与所研究的题目具有本质性联系,且能确切反映实

验因素的效应。否则,不可能作出正确的判断。

例如,在评价一种治疗骨折药物疗效的临床试验研究中,研究者确定的主要疗效指标为“骨密度”、“骨钙素”、“桡骨缺损处周长”等。显然,选择这些指标来评价药物对骨折的治疗效果既不敏感又不特异,失去了“针对性”,这些指标用来反映患者“骨质疏松状况”比较合适。那么,哪些指标用来反映骨折愈合效果合适呢?根据基本常识和骨科专业知识得知,应选择“骨形态”、“X线片”、“病理检查结果”和“功能恢复情况”等。

3.2 指标的观察 如果实验效应的观察带有偏性,则会影响结果的比较和分析。为了消除或最大限度地减少这种偏性,在实验设计时常采用盲法(blind method)^[1]。单盲法(single blind method):仅受试对象不知道自己被分在哪一组。双盲法(double blind method):受试对象和实验执行者均不知道受试对象被分在哪一组。三盲法(triple blind method):受试对象、实验执行者和统计分析人员三者均不知道受试对象被分在哪一组。双盲法和三盲法在临床试验中应用非常广泛。

3.3 指标的标准化 指标的标准化是指指标的采集方法、部位、时间、保存、运输、测定时间、测定方法、测定条件等方面的固定化。不进行标准化,这些因素会干扰研究进程,或者成为混杂因素,使指标的测定不稳定,影响研究结果。

3.4 指标的先进性 是指要选择高、精、尖、新的指标。课题的先进性一般由指标的先进性体现出来,先进指标更能深入地反映所研究的问题,揭示事物的本质。

参考文献

[1] 张学利,王善金,王云力,等.锚定法单开门椎管成形术对术后轴性症状和颈椎曲度影响的病例对照研究.中国骨伤,2008,21(10):759-761.
 [2] 胡良平.医学统计实战练习.北京:军事医学科学出版社,2007.2-3.
 [3] 佟海英,胡素敏,周鹏,等.悬吊模拟失重及解悬吊对大鼠骨密度及生物力学的影响.中国骨伤,2008,21(4):276-278.
 [4] 康甫,汪青春,叶勇光.旋转扳法和穴位推拿治疗青年颈性眩晕的随机对照实验.中国骨伤,2008,21(4):270-272.
 [5] 方弘伟,李建军,吴增晖,等.前下后上型齿突骨折螺钉内固定的生物力学研究.中国骨伤,2008,21(4):278-281.
 [6] 胡良平.统计学三型理论在实验设计中的应用.北京:人民军医出版社,2006.20-21.
 [7] 吴志宏,包飞.电针治疗膝关节炎临床对照实验.中国骨伤,2008,21(3):170-172.
 [8] 翟羽东,王寅,李宏兴,等.微创路径椎间盘髓核摘除术治疗腰椎间盘突出症.中国骨伤,2008,21(6):463-464.
 [9] 沈毅,严日树.医学统计学.上海:上海医科大学出版社,1999.86-88.
 [10] 高红兵,吴涛.关节外侧小切口治疗肱骨髁上骨折.中国骨伤,2008,21(2):113-114.
 [11] 方积乾.卫生统计学.第5版.北京:人民卫生出版社,2003.128-129.

(收稿日期:2009-01-15 本文编辑:李为农)