

(27) 87-91

医学实验中常见设计类型的辨析及统计方法的合理选用*

R311

胡良平

(军事医学科学院情报研究所医学统计教研室,北京 100850)

摘要 目的:帮助医学研究工作者提高对常见统计应用错误的识别能力,提高对统计学的正确利用率。**方法:**明确阐述各种常见医学实验设计类型的异同点和应用场合,揭示常见统计应用错误的实质和危害性。**结果:**给出不适用 t 检验取代方差分析的充足理由和正确运用统计学的基本要领。**结论:**要提高实际工作者的统计学应用水平,需要经常对其进行统计培训,积极引导他们与统计学工作者开展科研协作。

关键词: 单因素 K 水平设计; 析因设计; 重复测量设计; t 检验; 方差分析

医学实验、统计辨析、设计类型

中图分类号: R311

文献标识码: B

文章编号: 1000-6834(2000)01-0087-05

在医学实验研究中,为了考察多个处理因素的综合作用,同时为了有效地控制非处理因素的干扰和影响,需要借助多因素实验设计方法。由于实验中所涉及的因素(包括处理因素和重要的非处理因素,后者简称为区组因素)的个数和水平数不同,因素之间的组合方式不同等原因,需要许多不同的实验设计类型与之相对应。不同的设计类型,所能回答的问题是不同,对收集资料的要求也不同。因而,对实验资料的表达、描述、统计分析,甚至对结果的解释也是大有区别的。然而,从一些医学期刊的学术论文中却反映出一些令人担忧并带有相当普遍性的问题:即一律用算术均数加减标准差的方式表达定量的观测结果,一律用 t 检验处理观测结果为定量的实验数据,一律用 χ^2 检验处理观测结果为定性的实验数据^[1,2]。

应当清楚地意识到:用表达正态分布资料的方法表达明显呈偏态分布的资料,是对实验资料真实性的一种严重歪曲;不探究实验资料所取自的实验设计类型,就用分析单因素二水平设计资料的 t 检验进行处理,是对定量资料的肢解和破坏,是对因素之间客观存在的交互作用的无视或回避;不顾定性变量的有序性,盲目套用与有序性毫无关系的 χ^2 检验处理单向有序和双向有序的列联表资料,是对定性资料的浪费和糟蹋。误用和滥用统计学,并非仅仅是无关大局的一般性学术问题,而常常导致实验设计方面严重出错和对实验数据的错误解释,从而得出令人怀疑、甚至是错误的结论。误用和滥用统计学,是降低医学科研工作科学性的大敌,千万不可掉以轻心^[3]。

1 常见医学实验设计类型的辨析方法

1.1 常见医学实验设计类型简介^[4]

医学实验设计类型有几十种,其中常见的有:配对设计、成组设计、单因素 K 水平设计($K \geq 3$)、配伍组设计、拉丁方设计、交叉设计、析因设计、正交设计和具有重复测量的设计。从是否便于考察因素之间交互作用的角度看,前六种设计都不便考察交互作用,后三种设计是可以考察交互作用的。从同时考察因素的个数多少角度看,前三种设计都属于单因素设计,配伍组设计属于二因素设计、拉丁方设计、交叉设计都属于三因素设计,而后三种设计即可以用于二因素设计,又可以用于多因素设计。由于配伍组设计、拉丁方设计和交叉设计都不便考察交互作用,故最适合用于安排只含一个处理因素,含一个或二个区组因素的实验研究场合。如果实验中同时涉及二个或二个以上处理因素,因素之间的交互作用往往又是不可忽视的,此时,就应当选用析因设计或正交设计。如果希望观察接受不同处理的几组受试对象某些定量观测指标随时间推移的动态变化趋势,需要在不同时间点上从同一个受试对象身上进行多次观测,这就是所谓的重复测量设计。医学研究工作者常根据专业上的需要,自发地运用了重复测量设计,但在收集资料时,无意识地将重复测自同一个受试对象的多个数据与测自其他受试对象的数据混杂在一起,误将重复测量设计当作具有独立重复实验的析因设计^[2]。造成了在数据收集、统计分析方法选择和计算结果等方面出现了一系列的错误。

* 收稿时间:1999年7月19日;修回时间:1999年10月8日

作者简介:胡良平(1955-),男,安徽人,副教授,硕士,从事医学实验设计、多元统计分析与SAS软件应用的开发研究

误。

1.2 常见医学实验设计类型的辨析

下面借助生理学、药理学等医学研究领域里的实例,介绍常见医学实验设计类型的辨析方法。

【例1】某人在研究三七皂甙 Rgl 对大鼠实验性血栓形成的影响和对正常血压及高血压大鼠血小板内游离钙水平影响时,收集到表1和表2资料,均误用 t 检验作了统计分析。试判断这二张表里的资料所对应的实验设计类型是什么。

Tab. 1 Effects of Rgl, Rbl and lysinipirinum(LNP) on thrombosis in rats($\bar{x} \pm s$)

Drug/(mg·kg)	Thrombosis wet weight/mg
NS	25.3 ± 3.9
Rgl 5	8.7 ± 1.2
Rbl 5	18.0 ± 3.3
LNP 9.04	8.2 ± 1.2

NS: Normal saline; Rbl and Rgl: Two components in ginsenosides; LNP: Lysinipirinum

Tab. 2 Effects of Rgl on platelet $[Ca^{2+}]_i$ under the influence of thrombin in normotensive and hypertensive rats($\bar{x} \pm s$)

Group	Rgl/ ($\mu\text{mol} \cdot \text{L}^{-1}$)	$[Ca^{2+}]_i$ (nmol/L)	
		Resting	Thrombin
Normotensive	0	167 ± 24	503 ± 62
	1	154 ± 26	255 ± 66
	10	141 ± 30	160 ± 33
	100	120 ± 26	140 ± 31
Hypertensive	0	230 ± 47	709 ± 65
	1	182 ± 38	312 ± 46
	10	140 ± 39	232 ± 31
	100	131 ± 32	157 ± 37

【分析与解答】表1中虽然涉及到药物和剂量两个方面,但一种药物只取了一种剂量,并不能将剂量视为一个独立的因素。因此,在表1中,只有药物一个因素,它有四个水平,生理盐水可被视为一种特殊的“药物”。显然,与表1资料所对应的实验设计类型应叫做单因素四水平设计。

表2资料中涉及到三个因素,即受试者的类型(分为正常血压者与高血压者)、接受药物 Rgl 的剂量(分为 0.1、10 和 100 $\mu\text{mol} \cdot \text{L}^{-1}$)和受试者被检测时所处的状态(静息与接受凝血酶刺激后)。本资料所对应的实验设计类型究竟是什么?这取决于在二种状态下是对各组中每一只大鼠分别进行了检测,还是对不同大鼠进行了检测。若属于前一种情况,

应叫做具有一个重复测量的三因素设计;若属于后一种情况,应叫做三因素析因设计。

【例2】某人在研究富硒螺旋藻(缩写成 SES)对 ^{60}Co γ 线胸部照射大鼠诱发肺炎的防治作用时,收集到表3和表4资料。文中明确交代各组动物均在照后 0.5、1、2 和 3 个月时活杀,检测有关指标的数值,但均误用 t 检验作了统计分析。试判断这二张表里的资料所对应的实验设计类型是什么。

Tab. 3 Changes of BALFC($\times 10^6$) measured in left lungs of rats at different time points by ^{60}Co γ irradiation of thorax in rats($\bar{x} \pm s$), ($n = 6$)

Group	Time(d):			
	15	30	60	90
NC	1.0 ± 0.3	2.4 ± 0.5	2.5 ± 1.8	2.4 ± 1.2
AI	11.0 ± 5.5	15.8 ± 9.1	6.4 ± 2.3	7.0 ± 5.6
SI	7.2 ± 2.8	6.2 ± 1.2	3.0 ± 0.8	2.7 ± 1.0

NC, AI and SI stand for normal group, merely irradiation group, and irradiation plus selenium group

Tab. 4 Values of IOD measured from the collagen mRNA in rats' pulmonary tissue affected by SES($\bar{x} \pm s$)

Group	$\alpha_1(\text{I})^a$	$\alpha_1(\text{III})^b$
NC	1.68 ± 0.25	1.38 ± 0.15
NI	2.17 ± 0.21	1.73 ± 0.08
SI	1.83 ± 0.09	1.40 ± 0.11

$\alpha_1(\text{I})$ and $\alpha_1(\text{III})$ are two types of collagen, 4.6 kb and 4.5 kb respectively; a and b are two time points, 0.5 and 2 in month after irradiation respectively, when the values were measured

【分析与解答】在表3资料中,虽然有不同的四个时间点,但对每一只大鼠来说,仅在一个时间点上被观测,故此实验设计不属于重复测量设计。表3中涉及处理和时间二个因素,二因素之间的12种水平搭配下都独立地做了6次重复实验,故与该资料对应的实验设计类型应叫做二因素析因设计。

在表4中,从实验设计的严谨性角度看,存在一点问题。因为表中实际涉及了三个因素,即处理、胶原类型和照射时间,但胶原类型与照射时间的效应之间产生了混杂。因为照后0.5月的大鼠用的是第一种胶原,而照后2月的大鼠用的是第二种胶原,若二者之间的差别有显著性意义,不知是因胶原的类型不同所致,还是照后时间不同所致。于是,表4中纵向二个条件只能被视为一个混杂因素的二个水平。在这种假定之下,与表4资料对应的实验设计类型也应叫做二因素析因设计。

【例3】某人用 6.5 Gy 照射小鼠后再用几种不

同的药物治疗,观察每只小鼠照后不同时间点上外周血白细胞数的动态变化情况,各处理组中的小鼠在不同时间点上反复被观测。资料如表5所示。原

作者误用 t 检验分析了该资料,问与此资料所对应的实验设计类型应当是什么?

Tab. 5 Changes of circulating WBC($\times 10^9/L$) in rats by 6.5 γ irradiation($\bar{x} \pm s, n=10$)

Days after irradiat [#]	WBC($\times 10^9/L$)			
	Control	rhIL-6	rhG-CSF	rhIL-6 + rhG-CSF
Before*	19.7 \pm 1.8	17.3 \pm 3.2	17.9 \pm 3.8	18.3 \pm 3.7
7	1.8 \pm 0.7	1.9 \pm 0.8	2.0 \pm 1.0	2.2 \pm 1.5
10	1.8 \pm 0.9	1.7 \pm 0.8	1.7 \pm 0.6	1.3 \pm 0.6
14	1.9 \pm 0.6	2.9 \pm 1.0	2.7 \pm 1.1	2.5 \pm 0.9
17	2.8 \pm 1.0	3.3 \pm 0.9	4.6 \pm 1.2	4.5 \pm 1.1
21	7.6 \pm 2.0	6.6 \pm 2.4	6.0 \pm 3.0	6.9 \pm 4.2
23	9.4 \pm 3.8	9.7 \pm 4.2	9.7 \pm 3.9	8.7 \pm 5.6

irradiat[#] means irradiation; before* means before irradiation. rhIL-6, 200uKD; rhG-CSF, 125 uKD

【分析与解答】在表5中,究竟涉及了几个处理因素?很多人都认为它涉及了二个因素,一个是药物种类,另一个是照后时间。其实,它涉及了三个处理因素,除照后时间外,纵向所列的四组是各有二水平的二个因素的四种组合,并非是一个药物因素的四个水平,仔细观察不难看出,这个组合因素的四个水平就是由 rhIL-6 不用与用, rhG-CSF 不用与用的四种组合。将这四种组合视为四个实验条件,每个实验条件下的10只小鼠在7个不同的时间点上被重复观测 WBC 的值,故与表5对应的实验设计类

型应叫做具有一个重复测量的三因素设计。若将表5资料用表6的形式表达,则容易辨别出它所从属的设计类型。

除了上述已经接触到的重复测量设计形式外,根据整个实验所涉及的因素的多少以及在重复测量方向上所涉及到的因素数目还有具有二个重复测量的二因素设计和具有二个重复测量的三因素、四因素设计等,其设计和统计分析方法参见有关文献^[4]。

Tab. 6 Changes of circulating WBC($\times 10^9/L$) in rats at different time points by 6.5 γ irradiation($\bar{x} \pm s, n=10$)

Days after irradiat [#]	WBC($\times 10^9/L$)			
TREATMENT:	A ₀ , B ₀	A ₁ , B ₀	A ₀ , B ₁	A ₁ , B ₁
Before*	19.7 \pm 1.8	17.3 \pm 3.2	17.9 \pm 3.8	18.3 \pm 3.7
7	1.8 \pm 0.7	1.9 \pm 0.8	2.0 \pm 1.0	2.2 \pm 1.5
10	1.8 \pm 0.9	1.7 \pm 0.8	1.7 \pm 0.6	1.3 \pm 0.6
14	1.9 \pm 0.6	2.9 \pm 1.0	2.7 \pm 1.1	2.5 \pm 0.9
17	2.8 \pm 1.0	3.3 \pm 0.9	4.6 \pm 1.2	4.5 \pm 1.1
21	7.6 \pm 2.0	6.6 \pm 2.4	6.0 \pm 3.0	6.9 \pm 4.2
23	9.4 \pm 3.8	9.7 \pm 4.2	9.7 \pm 3.9	8.7 \pm 5.6

Irradiat[#] means irradiation; Before* means before irradiation. A₀ and A₁ stand for not using and using rhIL-6; B₀ and B₁ stand for not using and using rhG-CSF

2 常见医学实验设计类型资料的统计分析方法

2.1 为什么不能用 t 检验取代方差分析

t 检验每次只能比较二个平均数,因此,它仅适用于单因素一、二水平的实验设计类型。当设计类型为单因素多水平或多因素设计时,若仍用 t 检验分析资料,是不妥当的。其理由有以下五点:其一,

对原始资料的利用率很低,每次只能用到全部实验数据的几分之一;其二,破坏了原先的整体设计,将本属于多因素析因设计或重复测量设计割裂成多个单因素二水平的设计;其三,犯假阳性错误的概率大大增加,因为 t 检验只能将每次比较时犯假阳性错误的概率控制在规定的显著性水平上(即事先给定的 α 值),而多次 t 检验之后犯假阳性错误的总概率将随着比较的次数 n 的增大而趋向于1;其四,当实

验中同时涉及二个或二个以上处理因素时,因素之间往往存在不可忽视的交互作用,而 t 检验割裂了因素之间的内在联系,无法考察交互作用是否具有显著性意义;其五,由于不同组均数之间比较的 t 检验所用的标准误(即公式的分母)在不断改变, t 统计量的自由度小(因为资料的利用率低),所以,多次运用 t 检验的过程中,评价标准不统一。综上所述,充分说明用 t 检验处理非单因素一、二水平的设计资料所得结论的可靠性差^[2]。

应当明确的是:特定的实验设计类型,有相应的方差分析方法。方差分析之后,若某因素或交互作用具有显著性意义,又有相应的两两比较方法。这些方法可以克服 t 检验的上述不足。

2.2 方差分析的前提条件

不同的设计类型,需要运用相应的方差分析方法。但严格地说,方差分析也有其应用条件。与 t 检验相同,它也要求资料满足正态性(即各组数据分别服从正态分布)和方差齐性(即任何一个因素的多个水平所对应的多个总体方差应相等)。如果应用条件不满足,要么找到合适的变量变换方法,使变换后的数据满足条件,再对变换后数据进行方差分析;要么找到相应的非参数检验方法。

2.3 常见方差分析的种类及其合理选用要领

概括地说,方差分析方法有几十种。因为存在几十种不同的实验设计类型,而每一种实验设计类型,有与其对应的方差分析方法。分析的思路是基本相同的,只是具体的计算公式(即统计模型)是有区别的。

对于表 1 资料,如果四组资料都分别满足正态性、四组资料满足方差齐性,则应选用单因素四水平设计资料的方差分析。若方差分析的结果为四组均数之间的差别有显著或非常显著性意义($P < 0.05$ 或 $P < 0.01$),尚需作多个均数之间的两两比较。若四组中任何两组之间都需要比较,有多种方法,其中较常用的方法叫 q 检验;若三个药物组之间不需比较,而三个药物组都需与生理盐水组比较,则应选用 DUNNETT 的 t 检验。后者与通常的 t 检验(称为 STUDENT'S t -test)是有很大的区别的,欲知详情,请看有关文献^[4,5]。

对于表 2 资料,如果原始资料满足正态性和方差齐性,根据它是属于三因素析因设计还是属于具有一个重复测量的三因素设计,就应分别选用相应设计资料的方差分析方法。通过三因素析因设计资

料的方差分析,可以分析出三个因素各自的效应大小以及它们之间各级交互作用(包括任何二个因素之间和三个因素之间的交互作用)的效应大小。当交互作用具有显著性意义时,还可对其作进一步分解,以便发现因素之间的最佳水平组合。通过具有一个重复测量的三因素设计的方差分析,可以有效地排除个体差异和相邻时间点上观测值之间的高度相关性对结果的干扰和影响,比运用一般的方差分析方法能更合理、更有效地揭示因素及其交互作用的效应大小。

若资料满足正态性和方差齐性,对于表 3 和表 4 资料,应选用二因素析因设计资料的方差分析予以处理。对于表 6 资料,则需要运用具有一个重复测量的三因素设计资料的方差分析。

3 讨论与建议

因素的数目、因素之间的组合方式以及实验的具体实施方式等是判断实验设计类型的重要参考依据。当实验中涉及时间因素时,要特别注意区分何时属于重复测量,何时属于独立重复实验。一旦将二者混淆,就会导致误用统计模型,极易得出错误结论的后果。当实验中涉及到多个处理因素的水平组合时,人们常习惯将全部组合冠以“组别”字样(如表 5 中的“组别”),并把“组别”视为一个处理因素,进而,又把这个所谓的具有多个水平的处理因素拆成多个单因素二水平的设计,反复运用 t 检验进行分析。这是分析定量资料中最常见的一种错误,应当引起人们的足够重视。

对于重复测量设计,要事先编制好收集实验数据的表格,确保重复测自同一个受试者的 K 个数据写在同一行上,以便从总离均差平方和中分解出个体差异的影响,也有利于消除时间点之间相关性不等的影响^[2]。

在医学科研中,要解决的问题往往比较复杂,经常要运用多因素实验设计方法和相应的统计分析方法。这些方法掌握起来并非轻而易举,而实际工作者往往没有足够的时间通晓全部统计学知识。为了在医学科研工作与学术刊物中避免出现严重的统计学方面的错误,笔者建议应在下列三方面多作一些工作:其一,业务主管部门和学术期刊出版部门要首先提高对统计学作用的正确认识,采取切实可行的措施阻止有严重统计学问题的成果通过鉴定和论文公开发表;其二,有关领导和业务部门要努力创造条

件,使实际工作者经常有机会接受统计学知识的培训,提高他们应用统计学的水平和能力;其三,应大力倡导科研协作精神,做到人尽其才,物尽其用。实际工作者与统计工作者的密切合作,不仅可以达到事半功倍的效果,还可大大提高科研工作的科学性和严谨性。

利用国际上著名的统计分析系统(简称 SAS)^[5],可以有效地进行多因素实验设计和各种实验资料的统计分析,它可以帮助人们从依赖计算器和 t 检验的落后局面中解放出来。随着人类文明程度的不断提高,科研工作(包括医学期刊中论文的质量)的水平和质量也应不断提高,以适应时代的发展。

4 参考文献

- [1] 胡良平,刘惠刚.医学期刊中常见统计错误分析.中华医学写作杂志[J],1998;5(2):28.
- [2] 胡良平主编.医学统计应用错误的诊断与释疑[M].北京:军事医学科学出版社,1999:1-105.
- [3] 胡良平.医学科研与新药评价等工作中一个不可忽视的问题——轻视和误用统计学[J].军事医学科学院院刊,1996;20(3):202.
- [4] 胡良平主编.现代统计学与 SAS 应用[M].北京:军事医学科学出版社,1996:57-151.
- [5] SAS Institute Inc. SAS/STAT User's Guide, Release 6.03 Edition. North Carolina, USA: SAS Institute Inc. 1988. 549-637, 941-947.

THE DISCRIMINATIVE ANALYSIS OF THE COMMON DESIGN TYPES AND THE RATIONAL SELECTION OF THE STATISTICAL METHODS IN MEDICAL EXPERIMENTS

HU Liang-ping

(Department of Medical Statistics, Institute of Medical Information, Academy of Military Medical Sciences, Beijing 100850)

ABSTRACT

Aim: To improve the medical researchers' discriminative analysis of misused common statistics and the correct utilization ratio of statistics. **Methods:** To expound clearly the similarities and differences and the applicability of various common design types in medical experiments, and to reveal the cause and harm of these common errors in statistical application. **Results:** To present the sufficient reasons for irreplaceability of the analysis of variance by t -test and main points concerning correct application of statistics. **Conclusion:** Improvement of their utilization of statistics requires that medical researchers should receive constant training in statistics and guidance in order to cooperate with statisticians in the scientific and research work.

KEY WORDS: design of a single factor with k levels; factorial design; design with repeated measurements; t -test; analysis of variance

(上接第 47 页)

后进行实验;②2260~4300 m 组:在西宁饲养一个月后运到玛多县,饲养一个月后进行实验;③4300~2260 m 组:在玛多饲养一个月后再运回西宁,大白鼠再在西宁饲养一个月后进行实验。实验条件和气温与高原鼠兔相同。

(2)实验方法 实验动物用 10% 乌拉坦麻醉,由颈动脉取血备用。①血红蛋白电泳 取实验动物肝素抗凝血 1.5 ml,用 3000 r/min 离心 5 min,除去血浆加 2 倍体积的生理盐水。用同等的离心时间和转速,洗涤红细胞三次。之后加同等体积的双蒸水混匀,再离心。取血红蛋白液层 1 ml 备电泳分析。电泳采用聚丙烯酰胺凝胶,浓度为 7.5%,电极缓冲液为 Tris-甘氨酸(pH8.3)加样 15 μ l,电泳时间 2.5 h。用氨基黑 10B 显色,胶条放在 7% 醋酸液中保存。②红细胞压积(Hct) 用微量压积法测定。用目测法计数红细胞数(RBC)。用高铁氧化法测定血红蛋白含量(Hb)。③右心室/左心室重量比值(RVW/LVW) 称右心室重量及左心室加室间隔重量,用两者值比较,以反映右心室肥厚程度。

(3)统计学处理 三组大白鼠血液学结果采用单因素方差分析法,三组之间两两比较,进行 q 检验。二组高原鼠兔血液学结果采用 t 检验。

2 结果

(1)RBC, Hct, RVW/LVW 比值观察结果 2260~4300 m 组的 RBC, Hct, RVW/LVW 比值明显高于其它两组,均有显著性差异($P<0.01$)。Hb 4300~2260 m 组与其它两组相比有显著性差异($P<0.01$)。

玛多组高原鼠兔的 RBC 和 Hct 明显高于 4300~2260 m 组,两组相比有显著性差异($P<0.01$)。而玛多组高原鼠兔 Hb 和 RVW/LVW 比值虽高于 4300~2260 m 组,但二组相比无显著性差异($P<0.05$)。

(下转第 95 页)