**文章中的统计学方法描述**

统计学方法一直是科技期刊研究原著文章中的重要组成部分，在研究设计和样本量估算已完成的情况下，最重要的就是数据的统计学分析环节，但目前中国大多数医学期刊对统计学方法的描述还不够规范。现将统计学方法的描述要点作一整理，作者撰写研究原著类文章时，应遵循以下步骤描述统计学方法。

1.描述统计学软件信息：首先要对正文的统计学软件加以描述，需描述软件名称、来源厂家和版本。常用的软件包括SAS、STATA和SPSS等，还包括附带统计学功能的软件如GraphPad Prism 7等。

2.描述统计学指标：需说明研究中各描述性结果指标的表示方式。
2.1计量资料：正态分布资料包括均值(mean)、标准差(SD)和标准误(SEM)；非正态分布资料包括中位数(Median，M)和四分位数间距(P75-P25)。
2.2计数资料和等级资料：主要表示为构成比(如4/15，比重)和百分率(如63%，频率强度)，一般用n(%)表示。如想了解各种疾病在特定人群所占的比例，用构成比表示，如想了解哪一个年龄组发病率高，用百分率表示。
2.3效应量指标：主要包括比值比(odds ratio，OR)及相对危险度(Risk Ratio，RR)。95%置信区间(confidence interval，CI)，如OR(95%CI)=2.6(1.3-5.2)。

3.统计学分析方法
3.1计量资料的比较
3.1.1两组比较(正态分布)：
t检验：又称Student t检验，必须满足正态性，方差齐条件，主要包括两样本t检验(独立样本t检验及成组t检验)和配对样本t检验。
配对样本t检验的适用情况：
(1) 同一研究对象给予处理前、后比较(即自身配对)；
(2) 同一受试对象接受两种不同的处理；
(3) 配对的两个受试对象分别接受两种不同的处理；
(4) 同一对象的两个部位给予不同的处理。
3.1.2多组比较(正态分布)：
方差分析：两个及以上样本间均数的比较，采用成组和配伍设计，包括单因素、双因素、多因素、析因设计、重复测量方差分析(> 3个时间点数据比较)。在双因素、多因素、析因设计方差分析结果中一定要有主效应和交互效应的说明。注意，比较多组数据时，不能用t检验代替方差分析，主要原因是t检验破坏了原先的整体设计；出现假阳性错误的概率显著增加；t检验割裂了各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。应采用方差分析结合事后检验进行两两比较。
3.1.3非正态分布计量数据的非参数检验(秩和检验)：两组数据差异比较用Mann-Whitney U检验，多组数据差异比较用Kruskal-Wallis H法。
3.2 计数资料的比较
3.2.1两组比较：行χ2检验。
(1) n > 40并且所以理论数(T)大于5，则用Pearsonχ2检验；
(2) n > 40并且所以理论数(T)大于1并且至少存在一个理论数< 5，则用校正Pearsonχ2检验；
(3) n > 40或存在理论数(T) < 1，则用精确(Fisher)概率法；
(4) n < 40，用Fisher精确概率法。
(5)配对样本资料比较：可用配对四格表χ2检验；自身前后数据资料比较：McNemyerχ2检验。
3.2.2多组比较：行列表χ2检验。
3.3等级资料的比较：对组间等级资料的显著性检验采用非参数检验法。
3.3.1两组比较：成组设计资料用Wilcoxon两样本比较法比较，配对设计资料用符号秩和检验法比较。
3.3.2多组比较：成组设计用Kruskal-Wallis H法、Ridit法比较；多个样本两两比较用Nemenyi法比较；配伍组设计用Friedman秩和检验法比较。
3.4 相关和回归分析
3.4.1相关性分析：先作散点图，确定有线性趋势方可进行相关性分析。线性相关：Pearson相关性分析(正态分布)；秩相关：Spearman相关性分析(非正态分布，等级资料)。
3.4.2线性回归：包括因变量(结局)、自变量(因素) 和连续变量，数据需符合正态分布。简单线性回归：1个因变量，1自变量；多重线性回归：1个因变量，多个自变量。

3.4.3 Logistics回归：包括因变量(结局)和自变量(因素)。条件Logistics回归(配对，病例对照数据)，非条件Logistics回归(成组数据)。其中非条件Logistics回归包括2种，二元Logistic回归：是指因变量为二分类变量(是，否；患病，未患病)的回归分析；多元Logistic回归：是指因变量为有序或无序分类变量(轻、中、重；高中、低；优、良、中、差；A，B，C，D)的回归分析。
3.4.4 Cox回归：包括因变量(结局)和自变量(因素)，多用于生存分析。风险函数比(hazard ratio，HR)：是生存分析资料中用于估计因为某种因素的存在而使死亡/缓解/复发等风险改变的倍数。
3.4.5纳入回归模型的变量选择：单因素分析后，应当考虑应该将哪些自变量纳入回归模型进行多因素分析？一般情况下，建议纳入的变量有：
(1)单因素分析组间数据差异有显著性意义的变量(此时，最好将P值放宽一些，比如P < 0.1或P < 0.15等，避免漏掉一些重要因素)；
(2)单因素分析时，没有发现差异有显著性意义，但是临床上认为与因变量关系密切的自变量。

4 统计学方法描述举例：
4.1数据描述：①试验采用SPSS 22.0软件(美国IBM公司)进行统计学分析。②统计分析时先检查各研究中心完成例数、病例的脱落情况，然后进行两组患者入选时的人口统计学及基线各有关特征的分析，考察试验组和对照组之间的可比性。③计量资料采用均数、标准差、中位数、最小值和最大值、上下四分位数进行统计描述，计数资料采用例数和百分比进行统计描述。
4.2统计分析方法：①试验中各随访时间点两组间L2-4、股骨颈、Ward’s三角区骨密度值较基线的差值、血清钙、甲状旁腺素、骨钙素、白细胞介素10、白细胞介素6、肿瘤坏死因子α和胰岛素样生长因子1水平比较采用两样本t检验(数据正态分布)或Mann-Whitney U检验(数据非正态分布)。②组内各时间点上述数据差异比较采用重复测量方差分析及LSD事后检验比较。③两组不良反应发生率差异的比较采用Pearson χ2检验。④各组骨密度值指标、骨质疏松指标及炎性因子指标间的相关性分析采用Pearson相关分析法(数据正态分布)或Spearman相关分析法(数据非正态分布)。⑤检验水准(双侧)α = 0.05。