

· 论 著 ·

利用主成分分析法探索疫苗临床试验质量管理的构成

冉晓敏 梁庆香* 李安娜 曹芳芳 陈静华

柳州市疾病预防控制中心 广西柳州 545007

【摘要】目的 为提高疫苗临床试验现场质量管理提供科学依据。**方法** 通过调查分析、文献调研以及专家访谈等方法,初步确定疫苗临床试验现场质量管理的构成,然后将相关指标编制成问卷对疫苗临床试验研究者进行问卷调查,最后采用主成分分析法对数据进行分析。**结果** 通过特征根、陡坡图检验、方差贡献率和专业含义,提取了2个公因子,累计贡献率为49.898%,依次命名为“综合因子”、“准入因子”。**结论** 构建的疫苗临床试验现场质量管理模型,对疫苗临床试验现场质量管理工作有现实指导意义。

【关键词】 疫苗临床试验; 质量管理; 主成分分析

【中图分类号】 R186

【文献标识码】 A

【文章编号】 2095-7711 (2020) 12-001-03

【基金项目】 柳州市科学技术局资助项目 (2016G020220)

Using principal component analysis to explore the composition of quality management in vaccine clinical trials

RAN Xiaomin, LIANG Qingxiang, LI Anna, CAO Fangfang, CHEN Jinghua

Liuzhou Center for Disease Control and Prevention, Guangxi Liuzhou, 545007, China

【Abstract】Objective To provide scientific basis for improving the quality management of vaccine clinical trials. **Method** By means of investigation and analysis, literature research and expert interviews, the composition of quality management in clinical trials of vaccines was preliminarily determined, and then the relevant indicators were compiled into questionnaires to investigate the researchers of clinical trials of vaccines. Finally, the data were analyzed by principal component analysis. **Result** Through characteristic root, steep slope test, variance contribution rate and professional meaning, two common factors were extracted, with cumulative contribution rate of 49.898%, which were named “comprehensive factor” and “access factor” in turn. **Conclusion** The quality management model of vaccine clinical trials has practical guiding significance for the quality management of vaccine clinical trials.

【Key words】 Vaccine clinical trials; Quality management; Principal component analysis

随着大规模流行性传染性疾病的频繁爆发,人民群众的健康受到越来越大的威胁。控制传染病最有效的办法就是预防,疫苗是预防传染病的主要生物制品,因此,新疫苗的研制如火如荼,疫苗临床试验数量逐年增多,但其风险控制和质量管理难度增大^[1]。因此《中华人民共和国药品管理法》规定临床试验必须遵守《药物临床试验质量管理规范》(简称GCP),GCP是我国参照WHO和ICH-GCP,同时结合我国实际情况,于2003年首次修订发布^[2]。2013年10月31日颁布《疫苗临床试验质量管理指导原则(试行)》,该原则对疫苗临床试验的各环节提出质量管理要求^[3]。随后在2003年版GCP基础上,我国于2020年7月1日正式实施最新版GCP,新版GCP强调质量管理体系应覆盖临床试验全过程^[4]。但上述文件对质量管理没有具体的相关实施指南,且目前在疫苗临床试验中如何做好质量管理的尚不多,在疫苗临床试验质量管理方面,尚无统一的管理体系,难以正确指导和评价疫苗临床试验的质量管理工作,所以探索有效、可行的质量管理模型势在必行。

1 对象与方法

1.1 研究对象

对柳州市和下辖6个县疾病预防控制中心的疫苗临床试验研究者进行问卷调查,参与调查的研究者均为从事疫苗临床试验两年及两年以上的工作人员。共发放问卷211份,收回有效问卷203份,有效回收率为96.21%。

1.2 研究方法

1.2.1 问卷调查法

通过调查分析、文献调研以及专家访谈等方法,构建疫苗临床试验现场质量管理模型,然后编制成疫苗临床试验质量管理

问卷。问卷内容包括个人基本信息与影响质量管理工作的105项指标条目,在这105项指标条目中(由于篇幅有限,在本研究中不一一展开介绍各个条目。),Q1-Q26表示的是准入条件组,Q27-Q31表示的是质控体系组,Q32-Q88表示的是临床试验过程管理组,Q89-Q92表示的是数据管理组,Q93-Q98表示的是档案资料管理组,Q99表示的是统计分析组,Q100-Q105表示的是临床试验报告组,在这105项指标条目中,各条目采取likert等级进行重要性评分,1-5分别代表“一点也没有关系”、“关系较低”、“一般”、“重要”、“非常重要”。

1.2.2 统计学方法

利用EpiData 3.0软件建立数据库,然后采用双录入的方法录入数据。运用SPSS20.0软件对数据进行主成分分析(简称PCA)。

2 结果

2.1 调查对象的一般情况

调查对象总共有203名,男性占38.42%,女性占61.58%,年龄为50岁以下的占75.37%,职称为初级以上的占99.61%,受教育程度为本科以上的占46.30%。

2.2 数据的适用性分析

在进行PCA分析前,需先满足2项假设:假设1:观测变量是连续变量或有序分类变量,本研究观测变量符合假设1。假设2:变量之间存在线性相关关系。根据相关矩阵的结果(篇幅有限,不作展示)可知,组内各条目之间的相关系数基本大于0.3(由于Q1、Q20与其组内其它条目的相关系数基本小于0.3,故删掉。则质量管理条目为103条。),说明,提示满足假设2。本研究中,KMO统计值为0.885,Bartlett球形检验 $\chi^2=23586.328$, $P < 0.001$ 。综上所述,本研究的数据可采用PCA法进行分析。

2.3 PCA结果

1、本研究主要用4种方法来判定提取主成分的数量,分别

* 通讯作者:梁庆香(1972—),女,汉族,本科,副主任技师,主要从事质量管理及微生物检验。

是 1) 特征值大于 1: 采用 PCA 法进行因子提取, 按照特征值大于 1 的标准, 共提取了 18 个因子, 累计方差贡献率为 75.963%, 结果见表 2。2) 解释数据变异的比例: 既往研究认为提取的主成分至少应该解释 5-10% 的数据变异。根据这一标准, 我们认为本研究应该提取前 2 位主成分, 结果见表 1。3) 陡坡图检验: 在本研究中, 第 3 主成分之后的数据趋于平缓, 因此我们认为可以提取前 2 位主成分, 详见图 1。4) 解释能力判断: 提取后的主成分对

相应变量有解释能力。从表 2 可见, 前 2 位主成分的数据结构清晰, 没有存在两个主成分同时解释一个变量的情况。主成分 1 在 Q2-Q105 等多项指标上都有较大的载荷; 主成分 2 在 Q2-Q26 等多项指标上有较大的载荷, 因此可将主成分 1 命名为综合因子。因子 2 主要包含了准入条件组的指标条目, 将其命名为准入因子。

2、利用因子分析结果进行主成分分析。特征向量 = 因子负荷 / SQRT (特征值), 结果见表 3。

碎石图

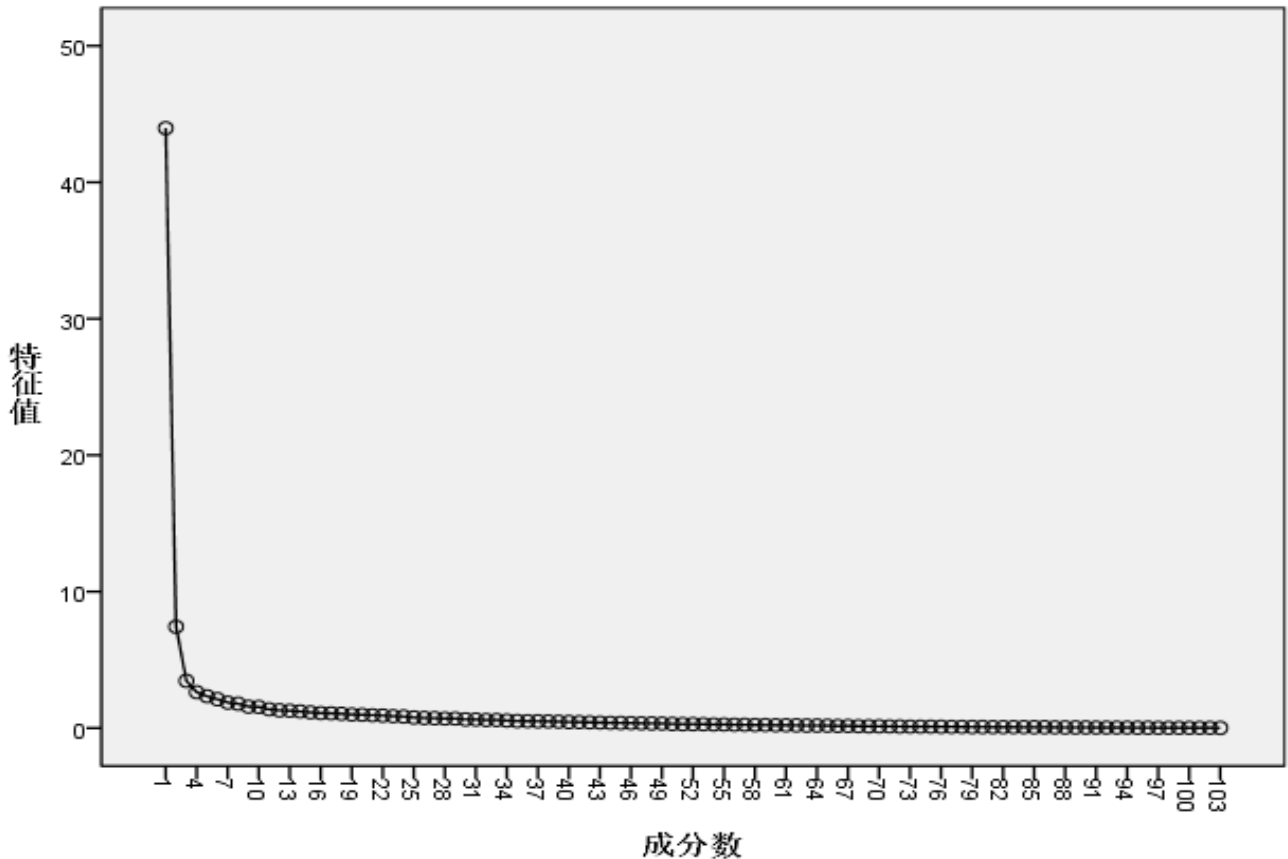


图-1: 陡坡图

表 1: 特征根和方差贡献表

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	43.965	42.685	42.685	43.965	42.685	42.685
2	7.430	7.214	49.898	7.430	7.214	49.898
3	3.458	3.357	53.256	3.458	3.357	53.256
4	2.627	2.551	55.806	2.627	2.551	55.806
5	2.333	2.265	58.072	2.333	2.265	58.072
6	2.117	2.055	60.127	2.117	2.055	60.127
7	1.868	1.814	61.941	1.868	1.814	61.941
8	1.788	1.736	63.676	1.788	1.736	63.676
9	1.571	1.525	65.201	1.571	1.525	65.201
10	1.556	1.511	66.712	1.556	1.511	66.712
11	1.372	1.332	68.044	1.372	1.332	68.044
12	1.300	1.262	69.306	1.300	1.262	69.306
13	1.268	1.231	70.537	1.268	1.231	70.537
14	1.215	1.180	71.717	1.215	1.180	71.717
15	1.151	1.117	72.834	1.151	1.117	72.834
16	1.099	1.067	73.902	1.099	1.067	73.902
17	1.082	1.050	74.952	1.082	1.050	74.952
18	1.041	1.011	75.963	1.041	1.011	75.963

3 讨论

本研究在调查对象的选取上, 男女比例约为 1.6, 多为中青年, 职称基本在初级以上, 本科以上的受教育程度约占一半, 其意见有一定的参考性。

进行全面评估的第一步是找出所选指标是否相互关联。如果是的话, 这样可以使评价结果偏向由于信息重叠^[5, 6]。为了解决该问题, 本研究尝试引入 PCA 来检查指标之间的相互关系, 并进一步实现更准确的评估。这个想法主要是基于该 PCA 可用来消除指标之间信息重叠的事实^[6, 7]。同时, 根据以往的研究^[8], 可通过特征根、陡坡图检验、方差贡献率和专业含义来综合判断提取主成分的数量。主成分分析结果提示, 本研究中前 18 个主成分的特征值大于 1, 但陡坡图分析提示应提取前 2 个主成分, 同时解释能力判断也提示提取前 2 个主成分比较符合研究实际需要。因此, 本研究最终提取前 2 个主成分, 分别反映疫苗临床试验质量管理的综合指标与准入指标。提取后的主成分累计解释 49.898% 的数据变异, 不到 60%, 但提取的每一个主成分对数据变异的解释比例都大于 5%, 具有一定的合理性。

在本研究中存在一些理论和实践上的贡献。首先, 它将多个变量归纳为某几项信息进行分析, 即降低数据结果的维度。其次, 它通过保留原始数据集的关键信息来提高 PCA 的适用性, 并可以更准确地表示评估结果。最后, 尽管这项研究提取后的主成分累

计解释只有 49.898%，不到 60%，但我们根据实际情况，在此基础上作进一步的完善。

参考文献：

[1] 高荣. 关于加强我国疫苗临床试验监管的思考(下). 中国新药杂志 2013.22.
 [2] 国家食品药品监督管理局. 药品临床试验质量管理规范(GCP), 2003.
 [3] 国家食品药品监督管理局. 《疫苗临床试验质量管理指导原则(试行)》, 2013.
 [4] 曾繁典. 药物临床试验管理规范与医学伦理[J]. 医药导报, 2020, 39(11):1466-1470.
 [5] Doorsamy W., Cronje W.A. A Method for Fault Detection on Synchronous Generators Using Modified Principal Component Analysis; Proceedings of the 2015 IEEE International Conference on Industrial Technology (ICIT); Seville, Spain. 17–19 March 2015; pp. 586–591.

表 2: 旋转前因子负荷矩阵

	成分		成分		成分			
	1	2	1	2	1	2		
Q88	0.818	-0.049	Q100	0.745	-0.074	Q50	0.638	-0.049
Q79	0.815	0.008	Q48	0.741	-0.069	Q63	0.631	-0.135
Q86	0.815	-0.039	Q82	0.739	0.025	Q27	0.629	-0.061
Q104	0.810	0.010	Q69	0.737	-0.104	Q35	0.628	-0.114
Q70	0.810	-0.026	Q44	0.735	-0.152	Q46	0.628	0.003
Q56	0.803	-0.092	Q37	0.734	-0.132	Q24	0.620	-0.045
Q97	0.799	0.116	Q99	0.733	0.066	Q34	0.607	-0.042
Q84	0.796	-0.065	Q71	0.730	-0.049	Q41	0.598	-0.204
Q60	0.796	-0.030	Q39	0.730	-0.149	Q51	0.585	-0.041
Q62	0.794	-0.041	Q23	0.726	-0.025	Q72	0.583	-0.065
Q81	0.785	-0.018	Q87	0.726	-0.078	Q22	0.564	0.019
Q96	0.784	0.102	Q98	0.723	0.109	Q36	0.560	-0.030
Q102	0.783	-0.039	Q95	0.722	0.054	Q26	0.546	0.015
Q57	0.782	-0.118	Q53	0.720	-0.060	Q25	0.543	0.067
Q61	0.782	-0.049	Q59	0.714	-0.133	Q38	0.516	-0.119
Q105	0.780	-0.067	Q74	0.705	0.009	Q94	0.484	0.120
Q78	0.774	0.033	Q52	0.705	0.022	Q40	0.470	-0.252
Q64	0.768	-0.038	Q90	0.704	-0.168	Q19	0.295	0.719
Q55	0.768	-0.124	Q66	0.704	-0.191	Q5	0.168	0.681
Q65	0.767	-0.020	Q89	0.703	-0.058	Q16	0.288	0.675
Q103	0.767	0.008	Q76	0.699	-0.115	Q12	0.150	0.673
Q80	0.761	0.138	Q32	0.699	0.015	Q18	0.225	0.659
Q93	0.760	0.077	Q28	0.698	-0.014	Q6	0.247	0.652
Q58	0.759	-0.072	Q45	0.698	-0.071	Q13	0.233	0.644
Q91	0.757	-0.078	Q42	0.694	-0.189	Q7	0.267	0.643
Q49	0.756	-0.058	Q68	0.688	-0.126	Q17	0.178	0.635
Q77	0.754	0.017	Q73	0.686	-0.045	Q9	0.227	0.630
Q67	0.752	0.002	Q47	0.672	-0.037	Q11	0.175	0.627
Q101	0.751	-0.016	Q29	0.665	0.044	Q3	0.201	0.623
Q30	0.751	-0.012	Q83	0.664	0.016	Q15	0.271	0.617
Q75	0.750	-0.051	Q54	0.656	0.030	Q8	0.252	0.560
Q85	0.748	-0.086	Q31	0.652	0.015	Q10	0.113	0.550
Q43	0.748	-0.154	Q33	0.645	-0.016	Q2	0.116	0.523
Q92	0.746	-0.019	Q21	0.644	0.045	Q4	0.181	0.449
						Q14	0.223	0.418

[6] Shirali G.A., Shekari M., Angali K.A. Quantitative assessment of resilience safety culture using principal components analysis and numerical taxonomy: A case study in a petrochemical plant. J. Loss Prev. Process Ind. 2016;40:277–284.

[7] Jolliffe I.T. Principal Component Analysis. 2nd ed. Springer; New York, NY, USA: 2002.

[8] 商立群, 王守鹏. 改进主成分分析法在火电机组综合评价中的应用[J]. 电网技术, 2014, 38(07):1928-1933.

表 3: 特征向量矩阵

	成份		成份		成份			
	1	2	1	2	1	2		
Q88	0.123	-0.018	Q100	0.112	-0.027	Q50	0.096	-0.018
Q79	0.123	0.003	Q48	0.112	-0.025	Q63	0.095	-0.049
Q86	0.123	-0.014	Q82	0.111	0.009	Q27	0.095	-0.023
Q104	0.122	0.004	Q69	0.111	-0.038	Q35	0.095	-0.042
Q70	0.122	-0.009	Q44	0.111	-0.056	Q46	0.095	0.001
Q56	0.121	-0.034	Q37	0.111	-0.048	Q24	0.094	-0.017
Q97	0.121	0.043	Q99	0.111	0.024	Q34	0.092	-0.015
Q84	0.120	-0.024	Q71	0.110	-0.018	Q41	0.090	-0.075
Q60	0.120	-0.011	Q39	0.110	-0.055	Q51	0.088	-0.015
Q62	0.120	-0.015	Q23	0.109	-0.009	Q72	0.088	-0.024
Q81	0.118	-0.007	Q87	0.109	-0.029	Q22	0.085	0.007
Q96	0.118	0.037	Q98	0.109	0.040	Q36	0.084	-0.011
Q102	0.118	-0.014	Q95	0.109	0.020	Q26	0.082	0.006
Q57	0.118	-0.043	Q53	0.109	-0.022	Q25	0.082	0.024
Q61	0.118	-0.018	Q59	0.108	-0.049	Q38	0.078	-0.044
Q105	0.118	-0.025	Q74	0.106	0.003	Q94	0.073	0.044
Q78	0.117	0.012	Q52	0.106	0.008	Q40	0.071	-0.092
Q64	0.116	-0.014	Q90	0.106	-0.062	Q19	0.045	0.264
Q55	0.116	-0.046	Q66	0.106	-0.070	Q5	0.025	0.250
Q65	0.116	-0.008	Q89	0.106	-0.021	Q16	0.044	0.247
Q103	0.116	0.003	Q76	0.105	-0.042	Q12	0.023	0.247
Q80	0.115	0.051	Q32	0.105	0.005	Q18	0.034	0.242
Q93	0.115	0.028	Q28	0.105	-0.005	Q6	0.037	0.239
Q58	0.115	-0.026	Q45	0.105	-0.026	Q13	0.035	0.236
Q91	0.114	-0.029	Q42	0.105	-0.069	Q7	0.040	0.236
Q49	0.114	-0.021	Q68	0.104	-0.046	Q17	0.027	0.233
Q77	0.114	0.006	Q73	0.103	-0.016	Q9	0.034	0.231
Q67	0.113	0.001	Q47	0.101	-0.013	Q11	0.026	0.230
Q101	0.113	-0.006	Q29	0.100	0.016	Q3	0.030	0.229
Q30	0.113	-0.005	Q83	0.100	0.006	Q15	0.041	0.226
Q75	0.113	-0.019	Q54	0.099	0.011	Q8	0.038	0.205
Q85	0.113	-0.032	Q31	0.098	0.005	Q10	0.017	0.202
Q43	0.113	-0.057	Q33	0.097	-0.006	Q2	0.017	0.192
Q92	0.112	-0.007	Q21	0.097	0.017	Q4	0.027	0.165
						Q14	0.034	0.154

根据表 3 可以得出主成分的表达式:

$$Y1 = 0.123 * Q88 + 0.123 * Q79 + 0.123 * Q86 + \dots + 0.071 * Q40 \quad (1)$$

$$Y2 = 0.264 * Q19 + 0.250 * Q5 + 0.247 * Q16 + \dots + 0.154 * Q14 \quad (2)$$

从上面的表达式可以看出, 第 1 主成分因子在质控体系组、临床试验过程管理组、数据管理组、档案资料管理组、统计分析组、临床试验报告组的各指标上有较高的载荷; 第 2 主成分因子在条件组的各指标上有较高的载荷。