

· 预防医学研究 ·

中国参与国际多中心双胞胎试验两种策略比较

石晓东^{1,2}, 成岗², 戴鲁燕³, 陈峰¹

¹南京医科大学公共卫生学院, 江苏 南京 211166; ²勃林格殷格翰(中国)投资有限公司, 上海 200040; ³和铂医药, 上海 201203

[摘要] 目的: 国际多中心临床试验(multi-regional clinical trials, MRCT)越来越多被各国药监机构所接受, 为评价试验结果的稳健性, 美国食品药品监督管理局(FDA)在某些治疗领域要求制药企业开展两个相同的国际多中心试验(又称双胞胎试验), 以确保试验结果的可重复性。本研究旨在评价中国以两种典型的策略加入国际多中心双胞胎临床试验的统计学效能。方法: 采用统计学模拟试验的方法, 比较两种策略的统计学性质, 以及中国人群与整体人群在疗效上的一致性概率。结果: 两种策略下按照标准1和标准2计算出的一致性概率来看, 并且仅考虑一致性概率的情况下, 策略A优策略B。结论: 给出了在不同情况下两种策略的优势与劣势, 并提供了相应策略。

[关键词] 临床试验; 国际多中心; 双胞胎试验; 一致性概率

[中图分类号] R

[文献标志码] A

[文章编号] 1007-4368(2019)02-295-05

doi:10.7655/NYDXBNS20190230

The comparison of two strategies for China as a center in multi-regional twin clinical studies

Shi Xiaodong^{1,2}, Cheng Gang², Dai Luyan³, Chen Feng¹

¹School of Public Health, NMU, Nanjing 211166; ²Boehringer Ingelheim China, Shanghai 200040; ³Harbour BioMed, Shanghai 201203, China

[Abstract] **Objective:** Multi-regional Clinical Trials (MRCTs) are more widely accepted by regulatory authorities from different countries to evaluate the robustness of clinical trial results. FDA requires pharmaceutical companies to conduct two identical multi-regional clinical trials in some therapeutical areas, which is named as twin studies, to ensure that the trials results are repeatable. The paper evaluates two strategies for China to participate MRCT in case twin studies are required. **Methods:** Data simulation is performed to evaluate these two strategies, and further assess the consistency in efficacy between China and overall population. **Results:** Accordind to the consistency probablinies calculate by standard 1 and standard 2, the strategy A is better than the strategy B in the two strategies. **Conclusion:** Strength and weakness of two strategies are illustrated under different scenarios, and provides corresponding solution and recommendation.

[Key words] clinical trial; MRCT; twin studies; consistency probability

[J Nanjing Med Univ, 2019, 39(02):295-299]

近年来,全球药品开发受到各大制药企业的广泛关注,国际多中心药物临床试验(multi-regional clinical trial, MRCT)定义为多个国家(区域、地区)的多个中心,或在某区域内不同国家的多个中心,按照同一临床试验方案同时开展的区域性临床试验,目的是使新药在多个国家或区域同时注册上市^[1-2]。MRCT可以维持试验设计在相同水平的科学严谨性,减少不必要的临床试验重复,获取全球的临床

试验数据,为各个地区药物批准提供更好的证据,加快新药的上市,有利于在多个地区申报注册新药^[3]。MRCT设计需要解决两个问题:①该药物在所有地域的“总疗效”是否存在?②在总疗效存在的条件下,其结论是否能推广到某个特定地区或各地域间疗效是否具有相似性^[4-5],对于判断总疗效是否能推广到某个特定地区,又与参加试验各地区监管机构的审批策略密切相关。2002年国家药品监

督管理局发布对国际多中心临床试验进行监管的具体规定,也就是自2002年起中国越来越多参与到国际多中心临床试验^[6-7]。

1 对象和方法

1.1 对象

双胞胎试验就是指两个设计完全相同或相似的临床试验,目的是验证药物的有效性以及结论的稳健性,确保试验的可重复性^[8],尽可能将试验的假阳性降到最低。

创新药物在全球开展国际多中心双胞胎临床试验,中国中心同时加入这个国际多中心双胞胎临床试验,可以实现该创新药物在全球的同步上市。本文将对中国受试者加入该类试验的两种策略进行统计学评价,以期同类研究提供参考。

中国中心参与全球双胞胎试验有两个典型的策略。策略A是所有中国受试者入组一个试验;策略B是中国受试者分别加入到两个试验。前者,中国亚组疗效结果与加入的该试验全球总人群进行比较以评估一致性;后者,中国亚组疗效结果分别与两个试验总体人群结果进行比较,同时,进一步比较两个试验中中国亚组合并之后与总人群合并之后结果的一致性。试验样本分配策略与统计分析方法的选择,需要考虑在整个试验成功(即总体人群得到疗效阳性结果)基础之上,结合适应证人群的特性,能够有效评估中国与总人群的一致性,并反映该结果在中国人群的可重复性和稳健性。

1.2 方法

对于区域或国家加入国际多中心试验,区域人群与总体人群的一致性,成为评估试验总人群疗效结果是否能够代表区域人群疗效的重要考量。常用的一致性评估统计方法主要包括两大类。一是基于某个特定区域人群的疗效与总体人群疗效的比较;二是基于每一个区域得到有利趋势的比较。

日本卫生劳动和社会保障部(Ministry of Health, Labor and Welfare, MHLW)在2007年9月提出了两种一致性评价标准。如果满足两个方法中的任意一个,就可以判定新药在日本地区与总人群疗效一致^[9]。

标准1: $P(D_J/D_{all} > k) \geq \gamma$

标准2: $P(D_1/0, \dots, D_n > 0) \geq \gamma$

其中J是关注的中心, D_J 表示该中心的疗效, D_{all} 表示总人群疗效, k表示该中心疗效占总人群疗效的比值,通常 $k=50%$; γ 常取80%。

标准1表示,如果中心J的疗效占总人群疗效50%的概率大于80%,则认为J中心的疗效与整体人群的疗效一致。标准2表示,所有中心的疗效估计都能呈现相同的趋势且疗效大于0的概率大于80%,则认为J中心的疗效与非J中心的疗效一致^[10]。

本文采用上述两种标准来评估双胞胎试验中中国人群相对于试验总人群在主要终点上的一致性,以评价不同策略的优劣。

2 结果

2.1 模拟试验目的

在中国受试者样本量固定的情况下,按照上述两种策略加入国际多中心双胞胎试验,计算这两种策略下中国人群与全球总人群疗效一致性概率,进而从统计学角度确定两种策略的优劣。

2.2 模拟试验设计及参数设置

不失一般性,假设试验是随机、双盲、安慰剂对照、平行组设计,两组样本量相同,主要结果指标是连续的且服从正态分布,且标准差相等,不妨设为1。

疗效定义为试验组与对照组主要结果指标之差值。记 D_C 为中国人群疗效, D_{NC} 为非中国人群疗效, D_{all} 为总人群疗效。 N_C 为中国受试者人数, N_{all} 为总人群受试者人数。各参数设置如表1,每种参数组合下模拟10 000次。模拟数据的产生和分析采用SAS 9.4编程完成。

表1 模拟参数设置

Table 1 Parameter settings in simulation

参数	取值
每组样本量(N)	100, 200
D_{all}	0.4, 0.6, 0.8
D_C/D_{all}	0.8, 1
N_C/N_{all}^1	10%, 20%, 30%, 40%, 50%
N_C/N_{all}^2	5%, 10%, 15%, 20%, 25%

N_C/N_{all}^1 : 中国中心按照策略A中国人群占总人群比例; N_C/N_{all}^2 : 中国中心按照策略B中国人群占总人群比例。

2.3 模拟试验结果

模拟试验结果如下表2和表3。

$D_C/D_{all} = 1$ 时,也就是假定中国人群的疗效与全球总人群疗效一致。 $D_C/D_{all} = 0.8$ 时,也就是中国人群的疗效不及全球总人群的疗效,仅是全球总人群的疗效的80%。

对于 N_C/N_{all}^1 为10%、20%、30%、40%和50%时,由于模拟是基于中国人群的样本量固定,同时对于策略B是中国中心分别加入两个试验,因此对应的

表2 不同参数组合时两种策略、两种标准下的把握度及一致性概率(D_c/D_{all}=1时)

Table 2 Power and probability of consistency under various parameter combinations in two strategies and standards (D_c/D_{all}=1)

每组样本量	D _{all}	策略A				策略B			
		N _c /N _{all} ¹	整个试验把握度(%)	标准1 [#]	标准2 [#]	N _c /N _{all} ²	整个试验把握度(%)	标准1 [#]	标准2 [#]
100	0.4	10%	80.42	0.682 8	0.814 5	5%	80.27	0.622 7	0.730 4
		20%	81.45	0.745 8	0.890 4	10%	80.22	0.670 5	0.802 1
		30%	80.49	0.811 7	0.932 2	15%	80.24	0.721 2	0.861 3
		40%	81.01	0.861 8	0.955 2	20%	80.29	0.751 7	0.893 3
		50%	79.92	0.896 2	0.954 2	25%	79.69	0.776 5	0.911 4
	0.6	10%	98.65	0.754 4	0.910 6	5%	98.67	0.682	0.826 8
		20%	98.63	0.850 4	0.971 4	10%	98.81	0.762 8	0.911 5
		30%	98.96	0.910 8	0.991 1	15%	98.67	0.816 2	0.951 9
		40%	98.65	0.945 5	0.997 2	20%	98.75	0.852 2	0.972 5
		50%	98.85	0.970 6	0.997 4	25%	98.68	0.882 0	0.983 4
	0.8	10%	100.00	0.826 2	0.962 5	5%	99.98	0.740 7	0.898 5
		20%	99.99	0.911 7	0.995 5	10%	99.99	0.818 6	0.961 4
		30%	99.99	0.961 5	0.999 2	15%	100.00	0.880 0	0.986 7
		40%	100.00	0.985 0	1.000 0	20%	99.99	0.916 7	0.994 0
		50%	100.00	0.994 4	0.999 9	25%	99.98	0.941 8	0.997 7
200	0.4	10%	97.81	0.750 1	0.903 4	5%	97.57	0.670 7	0.811 2
		20%	97.70	0.838 4	0.965 1	10%	97.68	0.741 1	0.895 7
		30%	97.70	0.896 3	0.985 8	15%	97.83	0.795 1	0.939 8
		40%	98.07	0.937 6	0.994 6	20%	97.91	0.838 9	0.966 3
		50%	97.81	0.962 0	0.996 3	25%	98.06	0.865 3	0.978 0
	0.6	10%	99.99	0.838 7	0.969 9	5%	100.00	0.757 1	0.914 2
		20%	99.99	0.930 4	0.997 0	10%	100.00	0.847 0	0.972 1
		30%	100.00	0.969 8	0.999 3	15%	100.00	0.889 7	0.988 5
		40%	99.99	0.985 8	1.000 0	20%	100.00	0.928 1	0.995 7
		50%	100.00	0.997 2	0.999 9	25%	100.00	0.953 6	0.998 4
	0.8	10%	100.00	0.907 9	0.994 3	5%	100.00	0.819 1	0.963 2
		20%	100.00	0.971 8	0.999 5	10%	100.00	0.910 1	0.993 5
		30%	100.00	0.993 5	1.000 0	15%	100.00	0.947 4	0.999 4
		40%	100.00	0.998 2	1.000 0	20%	100.00	0.973 0	0.999 7
		50%	100.00	1.000 0	1.000 0	25%	100.00	0.984 7	0.999 9

#:不同策略下不同标准下的一致性概率。

N_c/N_{all}²即为5%、10%、15%、20%和25%。由表2和表3的数据模拟结果,两种策略下按照标准1和标准2计算出的一致性概率来看,并且仅考虑一致性概率的情况下,策略A优于策略B。上述表中策略B下的一致性概率是在单个试验中中国人群与全球总人群的一致性概率,由于策略B下得两个试验是相互独立的,那么要使两个试验中中国人群疗效与全球总人群同时达到一致,则应是两个试验一致性概率的乘积,这样得出的一致性概率会更低。

上述模拟结果是基于试验药物有效情况下得出的,相反当H₀成立时也就是试验药物与对照药物没有差异,如果检出试验药物与对照药物差异有统计学意义,这是一类错误,应该控制在5%。利用模拟的数据计算得出的值为4.84%就是一类错误,控制得很好。

3 讨论

将中国纳入国际多中心临床试验已经成为国

表3 不同参数组合时两种策略、两种标准下的把握度及一致性概率(D_c/D_{all}=0.8时)

Table 3 Power and probability of consistency under various parameter combinations in two strategies and standards (D_c/D_{all}=0.8)

每组样本量	D _{all}	策略A				策略B			
		N _c /N _{all} ¹	整个试验把握度(%)	标准1 [#]	标准2 [#]	N _c /N _{all} ²	整个试验把握度(%)	标准1 [#]	标准2 [#]
100	0.4	10%	79.02	0.614 6	0.759 8	5%	80.03	0.578 9	0.686 9
		20%	77.12	0.671 6	0.839 1	10%	79.53	0.604 5	0.753 6
		30%	75.31	0.723 3	0.884 3	15%	77.78	0.640 8	0.806 8
		40%	73.79	0.764 4	0.904 3	20%	76.75	0.670 2	0.838 8
		50%	72.35	0.823 9	0.928 8	25%	76.36	0.697 7	0.864 9
	0.6	10%	98.58	0.665	0.856 4	5%	98.83	0.612 3	0.778 1
		20%	98.22	0.747 2	0.937 9	10%	98.51	0.661 9	0.858 8
		30%	97.78	0.807 2	0.968 6	15%	98.27	0.711 9	0.912 9
		40%	97.10	0.863 7	0.983 2	20%	98.35	0.747 7	0.933 5
		50%	96.88	0.911 0	0.990 4	25%	98.12	0.780 0	0.953 8
	0.8	10%	99.97	0.717 3	0.922 1	5%	99.99	0.652 9	0.838 7
		20%	100.00	0.811 8	0.975 9	10%	99.99	0.720 4	0.924 0
		30%	99.97	0.879 8	0.993 1	15%	99.98	0.777 5	0.958 8
		40%	99.97	0.927 1	0.997 8	20%	99.97	0.811 0	0.979 4
		50%	99.87	0.963 1	0.999 2	25%	99.97	0.843 5	0.989 6
200	0.4	10%	97.29	0.655 9	0.851 6	5%	97.73	0.610 7	0.758 4
		20%	97.05	0.731 8	0.923 8	10%	97.52	0.658 2	0.848 8
		30%	96.35	0.798 3	0.961 4	15%	97.00	0.691 6	0.891 0
		40%	95.60	0.851 3	0.977 1	20%	96.72	0.736 7	0.924 3
		50%	94.82	0.901 3	0.985 9	25%	96.59	0.760 9	0.944 7
	0.6	10%	100.00	0.726 7	0.935 8	5%	100.00	0.661 0	0.855 5
		20%	100.00	0.830 5	0.985 6	10%	100.00	0.734 9	0.938 3
		30%	99.99	0.893 7	0.995 9	15%	99.98	0.786 2	0.969 7
		40%	99.98	0.940 3	0.999 0	20%	99.99	0.821 5	0.982 5
		50%	99.01	0.952 9	0.995 8	25%	100.00	0.858 7	0.992 2
	0.8	10%	100.00	0.790 5	0.978 3	5%	100.00	0.712 0	0.925 5
		20%	100.00	0.894 3	0.997 8	10%	100.00	0.789 0	0.979 4
		30%	100.00	0.950 4	0.999 7	15%	100.00	0.859 7	0.994 9
		40%	100.00	0.978 3	1.000 0	20%	100.00	0.887 7	0.998 0
		50%	100.00	0.993 5	0.999 9	25%	100.00	0.918 2	0.999 0

#:不同策略下不同标准下的一致性概率。

内外制药企业创新药物研发的共同策略,以加快新药在中国上市,并推动中国临床研究国际化。国际多中心临床试验缩短了药物在不同区域和国家间的滞后,为医生和患者带来更多新的治疗手段和选择^[11-12]。

本研究中国中心加入国际多中心双胞胎试验的策略,在中国样本量固定的情况下,全部中国患者加入双胞胎试验中的单个试验(策略A),或按照MHLW标准1和标准2所得出的一致性概率比分别

加入两个试验(策略B)要高,从统计学角度来说,中国中心按照策略A加入全球双胞胎试验比策略B更具有优势。同时按照策略A,中国中心只加入全球双胞胎试验其中的一个临床试验,从临床运营来说减少了启动和关闭临床试验中心、伦理审批和提交等相关管理工作,提高临床运营效率。而中国中心按照策略B分别加入两个试验,由于所在各自的试验中中国患者样本量较小,可能会造成中国人群有效性指标的不一致。但是若某疾病在中国比较普

遍,中国有较多患者且比较容易招募,同时全球试验分配给中国中心的样本量较大,此时也可以考虑采取策略B。这样使中国亚组的疗效在保持与总体人群一致性的同时也可以使中国人群疗效在两个试验中的得到相互验证且可重复。

本研究数据模拟是基于连续正态分布变量,对于其他类型的终点指标,如二分类指标、生存时间等指标,本文的研究结果是否适用需要进一步研究。在已发表的文章中有对二分类终点资料的优效性设计进行了研究^[13]和正态资料下采用 α' 调整作为国际多中心临床试验目标区有效性研究^[14]。

本研究一致性评价标准采用MHLW提出的两个一致性评价标准,除了这两种标准外还有其他的一致性评价标准。例如,Carroll等^[15]提出的一种替代方法,Quan等^[16]提出的一种介于标准1和标准2的一种折中的方法。

本研究所得出的结论仅仅是从统计角度出发,然而从药物注册和监管要求层面来说,中国加入国际多中心双胞胎试验采取何种策略,制药企业应与监管机构应充分沟通交流,并且要得到相关监管部门的认可。

[参考文献]

- [1] 杨 焕,季双敏,高晨燕,等. ICH E17多地区临床试验指导原则的临床视角解读[J]. 中国新药杂志,2018,27(11):1238-1244
- [2] 黄 钦,王 骏,曲荣华,等. 探讨药物国际多中心临床试验设计的一种新方法 SGDDP[J]. 中国新药杂志,2015,24(7):721-724+744
- [3] 张晓方,武阳丰. 中国国际多中心临床试验的历史与现状[J]. 中国新药杂志,2018,27(11):1286-1289
- [4] Shih W J. Clinical trials for drug registrations in Asian-Pacific countries: proposal for a new paradigm from a statistical perspective. [J]. Controlled Clinical Trials, 2001, 22(4):357-366
- [5] Jun Shao, Shein-Chung Chow. Reproducibility probability in clinical trials. [J]. Statistics in Medicine, 2002, 21(12):1727-1742
- [6] 杨志敏,温宝书,史继峰,等. 国际多中心临床试验在中国申报和审评状况研究[J]. 中国医药工业杂志,2009,40(2):155-157
- [7] 袁 方. 我国开展国际多中心临床试验的现状与改进措施[J]. 医药导报,2007(01):100-102
- [8] John A Lewis. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline [J]. Statistics in Medicine, 1999, 18(5):1903-1942
- [9] Quan H, Zhao PL, Zhang J, et al. Sample size considerations for Japanese patients in a multi-regional trial based on MHLW guidance [J]. Pharmaceutical Statistics, 2010, 9(2):100-112
- [10] 柏建岭,陈梦锴,于 浩,等. 国际多中心临床试验中两种一致性评价方法的比较[J]. 中国临床药理学与治疗学,2017,22(09):1017-1021
- [11] 张晓方,黄 丹,王翔宇,等. 国际多中心临床试验监管指南研究报告[J]. 中国新药杂志,2017,26(17):2052-2058
- [12] 林 洁,魏永越,陈 峰,等. 中心效应及中心间样本不均衡对临床疗效评价的影响[J]. 中国临床药理学与治疗学,2012,17(4):428-432
- [13] 虞成凯,王 陵,戚洋洋,等. 基于 α 调整的国际多中心临床试验目标区二分类终点有效性研究的桥接方法 [J]. 中国卫生统计,2016,33(01):20-23
- [14] 尚 峰,虞成凯,杨 鹏. 正态资料下采用 α' 调整作为国际多中心临床试验目标区有效性研究的桥接方法 [J]. 中国临床药理学与治疗学,2018,23(1):59-64
- [15] Carroll, Kevin J, Le Maulf, et al. Japanese guideline on global clinical trials: statistical implications and alternative criteria for assessing consistency [J]. Drug Inform J, 2011, 45(5):657-667
- [16] Quan, Hui, Li, Mingyu, Chen, et al. Assessment of consistency of treatment effects in multiregional clinical trials [J]. Drug Inform J, 2010, 44(5):617-632

[收稿日期] 2018-09-29