

临界点分析法在处理临床研究 缺失数据中的应用

张盛婕 王杨 李卫

102300 北京, 中国医学科学院 北京协和医学院 国家心血管病中心医学统计部

通信作者: 李卫, Email: liwei@mrbc-nccd.com

DOI: 10.3760/cma.j.issn.0254-6450.2017.05.024

【摘要】 为应用可视化的临界点分析法处理临床研究缺失数据并探讨相关问题。枚举因缺失值导致的“所有”可能结局,并找出组间比较 P 值发生变化($P=0.05$ 为差异有统计学意义)的临界点,进一步计算 $P<0.05$ 的组合所占比例,判定研究结论的可靠性。临界点分析法可用于连续型或二分类结局的临床研究数据,并找到对应的临界点。在连续型结局实例中 $P<0.05$ 区域的比例为93.6%,否定零假设即试验成功的可靠性较高;二分类结局实例 $P<0.05$ 区域的比例为29.7%,否定零假设即试验成功的可靠性较低。临界点分析法可用于含缺失数据的临床研究,为结果判定提供可视化的判定依据。

【关键词】 临界点分析法; 临床研究; 缺失数据

基金项目:北京协和医学院青年基金暨中央高校基本科研业务费专项(3332015071)

Application of tipping-point analysis to address missing data in clinical studies Zhang Shengjie, Wang Yang, Li Wei

Medical Research and Biometrics Center, National Center for Cardiovascular Disease, Chinese Academy of Medical Science and Peking Union Medical College, Beijing 102300, China

Corresponding author: Li Wei, Email: liwei@mrbc-nccd.com

【Abstract】 To use a visualized method, tipping-point analysis to address missing data in clinical study and discuss related problems. All the possible outcomes caused by missing data were listed and the tipping points where P -values of hypothesis test changed at 0.05 significant level were found out, then the ratio of $P<0.05$ was calculated, reflecting the reliability of the study's result. Tipping-point analysis can be applied to both continuous and binary data to help find points where p -values are changed. The area of $P<0.05$ of continuous data is 93.6%, indicating that the reliability of success of the study is large; and the area of $P<0.05$ of binary data is 29.7%, reflecting that the reliability of success of the study is small. Tipping-point analysis, which provides a visualized evidence for decision making, is suitable for analyzing clinical studies with missing data.

【Key words】 Tipping-point analysis; Clinical study; Missing data

Fund program: Peking Union Medical College Youth Fund and the Fundamental Research Funds for the Central Universities (3332015071)

临床研究中由于研究对象病情恶化、死亡、失访等原因经常出现缺失数据导致结果发生偏倚^[1],甚至可能完全改变研究结论^[2-3]。近期美国国家研究委员会(National Research Council, NRC)提出了试验设计和实施阶段减少缺失数据的系统指南^[4-5],但缺失数据仍不可避免。常用于处理缺失数据的方法包括完整观测法(complete-case analysis)、填补法(imputation)、极大似然法(maximum likelihood estimation)和加权调整法(nonresponse weighting)。与以往传统方法不同,临界点分析法(tipping-point

analysis)是将因缺失值导致的“所有”可能结果通过枚举的方式列出,并逐一进行假设检验,找到使结论发生转变的组合,从而得到缺失数据存在情况下最全面的结果。2014年Liublińska和Rubin^[6]在美国食品药品监督管理局提出的临界点展示法(tipping-point display)基础上^[7],改进了结果可视化和连续型结局的处理方法,绘制临界点分析热图,并提出连续型结局缺失值均值取值范围的假设。为此本文旨在引入该方法,并利用临床试验实例分析和探讨该方法在实际应用中的问题,为临床研究者提

供一种简单易行处理缺失数据的方法。

基本原理

临界点(tipping point)^[7]定义为导致研究结果发生转变时,治疗组和对照组成功例数的差值(二分类结局)或均数的差值(连续型结局)。临界点分析法首先枚举因缺失值导致的“所有”可能结果,再进一步找到使组间比较的P值发生变化(与显著性水平比较)的组合, $P < 0.05$ 表明否定零假设,这些组合所占比例越大,表明该试验成功的可靠性越大。

1. 连续型结局:已知在某试验中,治疗组观测值的均数为 $\bar{y}_{obs}^T = \sum_{i: d_i^T=1} y_i / N_{obs}^T$, 对照组观测值的均数为 $\bar{y}_{obs}^C = \sum_{i: d_i^C=1} y_i / N_{obs}^C$; 假设治疗组缺失值的均数为 $g(Y_{mis}^T) = \bar{y}_{mis}^T = \sum_{i: d_i^T=0} y_i / N_{mis}^T$, 对照组缺失值的均数为 $g(Y_{mis}^C) = \bar{y}_{mis}^C = \sum_{i: d_i^C=0} y_i / N_{mis}^C$ 。由此可以求出在治疗组和对照组缺失值均数的不同组合下估计效应值 $\hat{\tau}$ (组间差异)

$$\hat{\tau} = \bar{y}^T - \bar{y}^C = \frac{\bar{y}_{obs}^T N_{obs}^T + g(Y_{mis}^T) N_{mis}^T}{N^T} - \frac{\bar{y}_{obs}^C N_{obs}^C + g(Y_{mis}^C) N_{mis}^C}{N^C}$$

由于不清楚实验组和对照组缺失值的变异程度,需要给出缺失数据的标准差假设。Liublinska 和 Rubin 等^[6]提出的标准差假设

$$s_T^2 = \frac{(K_{obs}^T - 1)s_{obs}^{T2} + \frac{K_{obs}^T K_{mis}^T}{K^T} (\bar{y}_{obs}^T - \bar{y}_{mis}^T)^2}{K_{obs}^T}$$

$$s_C^2 = \frac{(K_{obs}^C - 1)s_{obs}^{C2} + \frac{K_{obs}^C K_{mis}^C}{K^C} (\bar{y}_{obs}^C - \bar{y}_{mis}^C)^2}{K_{obs}^C}$$

式中 K_{obs}^T 、 K_{mis}^T 、 K^T 、 K_{obs}^C 、 K_{mis}^C 、 K^C 分别表示治疗组观测到结局数量、缺失数量、观测和缺失的总数及对照组观测到结局的数量、缺失数据量、观测和缺失的总数; s_{obs}^{T2} 、 s_{obs}^{C2} 分别是治疗组和对照组观测值(非缺失数据)的方差。

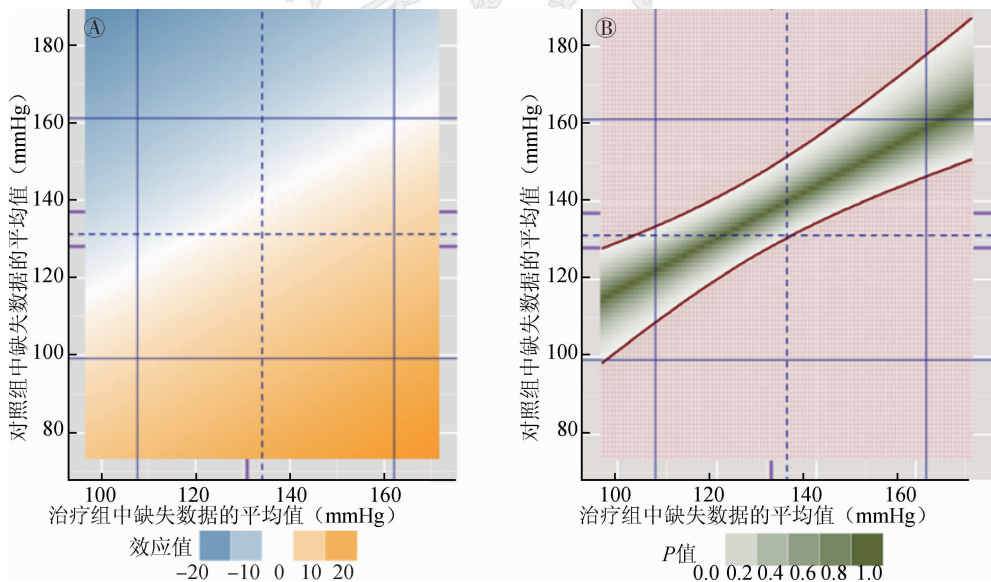
由此可进一步假设缺失值均数的取值范围,参考值为 $\bar{y}_{mis} \pm 1.5 \times s^2$ 。结合 Welch's t 检验公式,可求得每个组合的检验统计量 t,从而求出相应的组间比较 P 值

$$t = \frac{\hat{\tau}}{\sqrt{\frac{\hat{\sigma}_T^2}{N^T} + \frac{\hat{\sigma}_C^2}{N^C}}} = \frac{\bar{y}^T - \bar{y}^C}{\sqrt{\frac{\hat{\sigma}_T^2}{N^T} + \frac{\hat{\sigma}_C^2}{N^C}}}$$

$$\hat{f} = \frac{(s_1^2/N^T + s_0^2/N^C)^2}{(s_1^2/N^T)^2/N_{obs}^T + (s_0^2/N^C)^2/N_{obs}^C}$$

根据不同组合对应的估计效应值 $\hat{\tau}$ 和 P 值,可以画出反映 $\hat{\tau}$ 和 P 值大小的热图(图1)。

2. 二分类结局:与连续型结局类似,已知在某试验中,治疗组观测值的成功数为 \bar{y}_{obs}^T , 对照组观测值的成功数为 \bar{y}_{obs}^C ; 假设治疗组缺失值的成功数为 $g(Y_{mis}^T) = \bar{y}_{mis}^T N_{mis}^T$, 对照组缺失值的成功数为



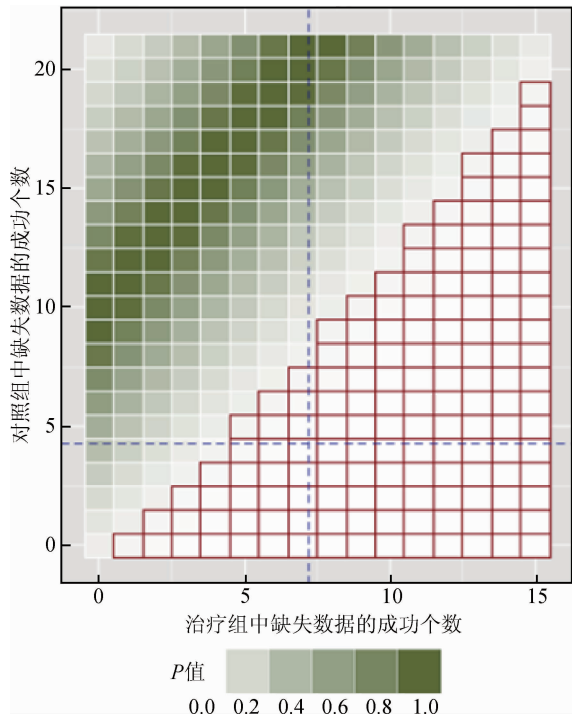
注:纵向和横向的蓝色虚线分别代表治疗组和对照组观测值的均数,蓝色实线代表治疗组和对照组观测值的取值范围,均为解读P值图时应重点关注的区域。A:橙色部分示 $\hat{\tau}$ 为正值,即治疗组优于对照组(高优指标)或治疗组劣于对照组(低优指标),蓝色部分示 $\hat{\tau}$ 为负值,即治疗组劣于对照组(高优指标)或治疗组优于对照组(低优指标); B:红色曲线示 $P=0.05$ 的边界

图1 某临床试验中两组患者血压的连续性临界点分析估计效应值 $\hat{\tau}$ 和 P 值

$g(Y_{mis}^C) = \bar{y}_{mis}^C N_{mis}^C$ 。由此可以求出在治疗组和对照组缺失值成功数的不同组合下的估计效应值 $\hat{\tau}$ (组间成功率差值)

$$\hat{\tau} = \bar{y}^T - \bar{y}^C = \frac{\bar{y}_{obs}^T N_{obs}^T + g(Y_{mis}^T)}{N^T} - \frac{\bar{y}_{obs}^C N_{obs}^C + g(Y_{mis}^C)}{N^C}$$

由治疗组总的成功数和失败数、对照组总的成功数和失败数所得四格表的 χ^2 检验可求得两组不同缺失值成功数组下的 P 值,从而画出估计效应值 $\hat{\tau}$ 和 P 值相应的热图(图 2),并可计算占全部组合情况的比例。



注:绿色区域颜色越深 P 值越大;红格区域代表 $P < 0.05$ 的情况
图 2 某临床试验中两组患者数据二分类临界点分析 P 值

3. 阳性结果比例(proportion of positive scenarios, POPS):其概念为否定零假设即 $P < 0.05$ 的组合数占所有组合数的比例(二分类结局),或否定零假设即 $P < 0.05$ 的区域面积占总面积的比值(连续型结局),可反映试验成功的可靠性,即 POPS 值越大,试验成功的可靠性越大。

4. 适用数据类型:适用各种类型的临床研究,包括随机对照临床试验、队列研究、横断面研究、病例对照研究等,只要存在 ≥ 2 个暴露因素,产生不同的暴露结局,即可应用该方法分析结局指标的缺失数据。但仅适用于处理结局指标含有缺失数据的情况,不可用于处理基线指标、研究因素为自变量的缺失数据。

实例分析

1. 连续型结局:为评价新一代药物洗脱支架治疗原发原位冠心病患者安全性和有效性,将原发原位冠心病患者作为受试对象,在植入支架后第 1、6、9、12、18 个月和第 2~5 年中定期进行临床随访。其中,第 9 个月造影随访的测量结果(晚期管腔丢失水平)是该试验的主要疗效指标(表 1),其比较类型为非劣效,界值设为 0.12 mm。

表 1 新一代药物洗脱支架治疗原发原位冠心病患者第 9 个月造影随访测量

分 组	试验组(n=245)	对照组(n=235)
缺失例数	31	32
缺失百分比(%)	0.126 5	0.136 2
观察管腔丢失水平(mm)		
均值	0.155 8	0.137 9
最小值	-0.38	-0.48
最大值	2.53	2.53

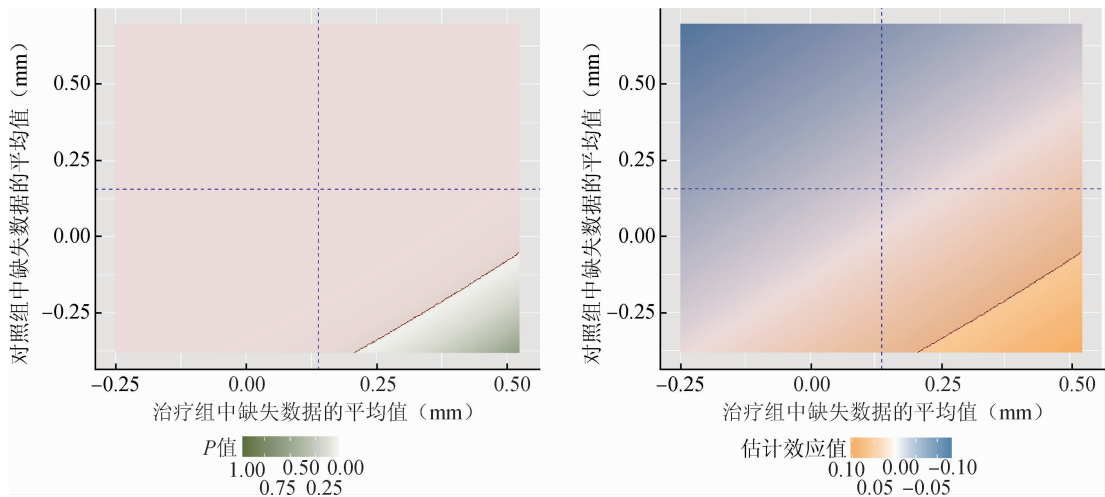
因造影随访为有创检查,通常支架试验中均有一定比例的患者缺少第 9 个月造影随访的结果,故使用临界点分析法对其缺失数据作分析图(图 3)。由于该试验的主要疗效指标为低优指标,故边界左上方代表“试验成功”,也即 $P < 0.05$ 否定零假设 $\hat{\tau} \geq \Delta$ 的区域。POPS 为 93.6% 可初步判定否定零假设,即实验成功的可靠性较高。

2. 二分类结局:为评价某医疗器械用于开颅术后颅骨骨瓣复位固定的有效性和安全性,以开颅手术且适合行自体骨瓣原位回植者为受试对象。术后 7(± 1) d、90(± 7) d 和 180(± 14) d 分别进行随访,主要终点指标为术后 180 d 根据骨瓣固定效果判断的有效率,结果见表 2。该试验的设计类型为非劣效,非劣效界值为 5%。

表 2 某医疗器械用于颅骨骨瓣复位固定术后 180 d 骨瓣固定效果

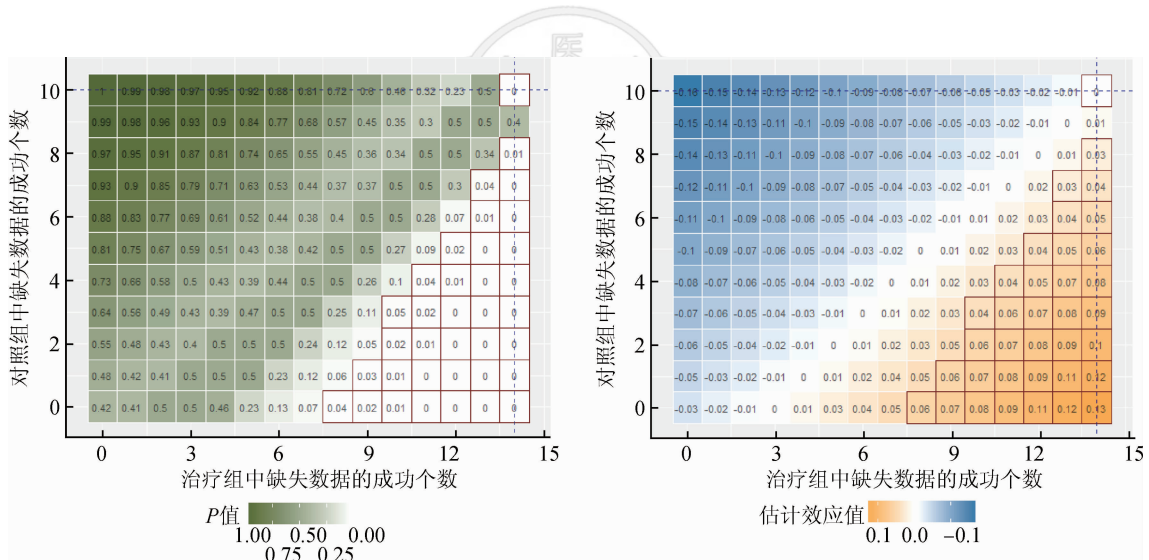
指 标	试验组(n=86)	对照组(n=77)
缺失例数	14	10
缺失百分比(%)	16.28	12.99
手术成功例数	72	67
手术失败例数	0	0

由临界点分析法可得图 4。由于该试验的主要疗效指标为高优指标,故边界右下方代表“试验成功”,也即 $P < 0.05$ 否定零假设 $\hat{\tau} \geq \Delta$ 的区域。POPS 为 29.7% 可初步认为否定零假设,即试验成功的可靠性相对较低。



注:左图标出估计效应值 $\hat{\tau}$ 趋于非劣效界值的边界,其左上方红色代表治疗组非劣于对照组,即 $P < 0.05$ 的区域,下方绿色代表 $P > 0.05$ 的区域。便于临界点分析结果展示,右图综合了左图的信息和估计效应值,橙色示估计效应值 $\hat{\tau} > 0$ 的组合区域,蓝色示估计效应值 $\hat{\tau} < 0$ 的组合区域,红色分界线对应于左图中估计效应值 $\hat{\tau}$ 趋于非劣效界值 Δ 的临界点构成的边界

图3 两组患者连续型结局临界点分析的P值(左)及估计效应值和P值(右)



注:左图红格代表 $P < 0.05$ 的组合,表示治疗组非劣于对照组,即 $P < 0.05$ 的区域;右图综合左图的信息和估计效应值,橙色示估计效应值 $\hat{\tau} > 0$,蓝色示估计效应值 $\hat{\tau} < 0$,红色阶梯形边界对应于左图中估计效应值 $\hat{\tau}$ 趋于非劣效界值 Δ 的临界点组合构成的边界

图4 两组患者二分类结局临界点分析的P值(左)及估计效应值和P值(右)

讨论

在本文连续型结局实例中, $P < 0.05$ 的区域占比较大(93.6%),否定零假设即试验成功的可靠性较高;在二分类结局实例中, $P < 0.05$ 的区域占比较小(29.7%),否定零假设即试验成功的可靠性较低。由此认为,临界点分析法可用于有缺失数据的连续型或二分类结局的数据,通过找到临界点并计算试验成功的情况占所有可能情况的比例,反映试验成功的可靠性,从而对含有缺失数据的临床研究结论作出全面分析和准确判断。

在使用临界点分析法时须注意,首先该方法只能用于分析结局变量的缺失数据,不能分析暴露因素的缺失数据,只要研究中涉及 ≥ 2 种暴露因素,产生不同的暴露结局,就可以分析结局变量的缺失数据。其次,该方法不同于传统的处理缺失数据方法,其目的并不是找到最优填补策略将数据补齐,而是列出所有可能的补齐情况,即通过研究结论所有情况的比例来反映结论的可靠性。

POPS是临界点分析量化的统一指标。本文引入了POPS的概念,即否定零假设的组合数占所有组合数的比例(二分类结局)或否定零假设的区域

面积占总面积的比值(连续型结局)。POPS反映了试验成功的可靠性,POPS越大,试验成功的可靠性越大。目前已可通过软件实现POPS的计算。

临界点分析法中“试验成功”的面积区域形状有时是条带状,有时是三角形或多边形,有时取右下角区域,有时取左上角区域,选择面积区域的依据是什么呢?实际上,“试验成功”区域主要取决于假设试验类型是优效性检验还是非劣效检验,以及主要疗效指标是高优指标还是低优指标。对于差异性检验,图形一般为条带状(如图1),右下角表示治疗组优于对照组,左上角表示对照组优于治疗组;对于优效性检验和非劣效检验,图形为多边形区域。对于高优指标,治疗组的值比对照组的值越大越好,“试验成功”区域应为右下角的区域(如图2和图4);对于低优指标,治疗组的值比对照组的值越小越好,“试验成功”区域应为左上角的区域(如图3)。

临界点分析法也存在一些影响POPS的因素,如连续型结局均值取值范围的确定、不同组合事件发生的概率的差别,均可能导致临界点分析结果产生一定的不确定性。此外,POPS的界值也尚无明确划分依据,还需在后续研究中完善。

综上所述,临界点分析法在处理临床研究缺失数据中与传统方法相比具有一定的优势,应受到相应的关注并在临床研究的缺失数据处理中推广使用。

志谢 国家心血管病中心流行病学研究部硕士研究生韩西坤为编写R包TippingPoint提出建议并协助修改编写程序
利益冲突 无

参 考 文 献

[1] Ware JH, Harrington D, Hunter DJ, et al. Missing data [J]. *N Engl J Med*, 2012, 367 (14) : 1353-1354. DOI: 10.1056/NEJMsm1210043.

[2] van der Heijde D, Fleischmann R, Wollenhaupt J, et al. Effect of different imputation approaches on the evaluation of radiographic progression in patients with psoriatic arthritis: results of the RAPID-PsA 24-week phase III double-blind randomised placebo- controlled study of certolizumab pegol [J]. *Ann Rheum Dis*, 2014, 73 (1) : 233-237. DOI: 10.1136/annrheumdis-2013-203697.

[3] Cuffe R, Barnett C, Granier C, et al. Missing CD4⁺ cell response in randomized clinical trials of maraviroc and dolutegravir [J]. *HIV Clin Trials*, 2015, 16 (5) : 170-177. DOI: 10.1179/1945577115Y.0000000003.

[4] Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials [J]. *N Engl J Med*, 2012, 367(14) : 1355-1360. DOI: 10.1056/NEJMSr1203730.

[5] Little RJ, Cohen ML, Dickersin K, et al. The design and conduct of clinical trials to limit missing data [J]. *Stat Med*, 2012, 31 (28):3433-3443. DOI:10.1002/sim.5519.

[6] Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial [J]. *Stat Med*, 2014, 33 (24) : 4170-4185. DOI: 10.1002/sim.6197.

[7] Yan X, Lee S, Li N. Missing data handling methods in medical device clinical trials [J]. *J Biopharm Stat*, 2009, 19 (6) : 1085-1098. DOI:10.1080/10543400903243009.

(收稿日期:2016-10-13)

(本文编辑:张林东)

中华流行病学杂志第七届编辑委员会通讯编委名单

(按姓氏汉语拼音排序)

- | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|
| 陈曦(湖南) | 党少农(陕西) | 窦丰满(四川) | 高婷(北京) | 高立冬(湖南) | 还锡萍(江苏) | 贾曼红(云南) |
| 金连梅(北京) | 荆春霞(广东) | 李琦(河北) | 李十月(湖北) | 李秀央(浙江) | 林玫(广西) | 林鹏(广东) |
| 刘莉(四川) | 刘玮(北京) | 刘爱忠(湖南) | 马家奇(北京) | 倪明健(新疆) | 欧剑鸣(福建) | 潘晓红(浙江) |
| 彭晓旻(北京) | 彭志行(江苏) | 任泽舫(广东) | 施国庆(北京) | 汤奋扬(江苏) | 田庆宝(河北) | 王丽(北京) |
| 王璐(北京) | 王金桃(山西) | 王丽敏(北京) | 王志萍(山东) | 武鸣(江苏) | 谢娟(天津) | 解恒革(海南) |
| 严卫丽(上海) | 阎丽静(北京) | 么鸿雁(北京) | 余运贤(浙江) | 张宏伟(上海) | 张茂俊(北京) | 张卫东(河南) |
| 郑莹(上海) | 郑素华(北京) | 周脉耕(北京) | 朱益民(浙江) | 祖荣强(江苏) | | |