RCT随机分组后组间基线不均衡的可能性大吗？

作者：陶立元 赵一鸣 来源：临床流行病学和循证医学

本文原地址：https://www.iikx.com/news/statistics/1444.html

在某RCT研究设计的讨论会上，研究者们打算对研究对象进行随机分组，但是又担心随机分组以后基线资料组间依然不均衡(即差异有统计学意义)，如病人的年龄、性别和疾病严重程度等。这时候有些人就开始纠结是采用完全随机分组好呢，还是采用分层随机分组好呢?假设该疾病的严重程度可以分为轻度、中度和重度3类，如果采用分层随机研究者需要在轻度患者中进行一次随机分组，分为A、B两组;再采用随机分组的方法将中度患者分为AB两组，同样去处理重度患者。这样的分层随机分组方法是否会好于完全随机分组呢?分层随机分组肯定会增加麻烦，但是这种麻烦的增加是否值得呢?先不说答案，咱们来模拟一组数据试试。

下面咱们就在R软件中做一下模拟，来看看结果会是什么样子的。此处定义完全随机分组为将研究对象随机分为例数相同的AB两组，分层随机分组也在同一层内将研究对象等分，同时再模拟一下不同样本量的研究对象是否会有差别。

1、完全随机分组和分层随机分组的结果

假设有600例研究对象，其中轻度、中度和重度患者各占300、200和100例。首先进行完全随机分组(grade表示严重程度等级)，重复分组5000次：

```
grade=c(rep(1, 300), rep(2, 200), rep(3, 100))

bg=rep(c(0,1), 300)

b=data.frame(rep(-1,5000))

for(i in 1:5000){

set.seed(i)

group=sample(bg, 600, rep=F)

a1= chisq.test(grade, group)

a2= fisher.test(grade,group)

b[i,1]=a1$p.value
```

b[i,2]=(b[i,1]<0.05)

b[i,3]=a2$p.value

b[i,4]=(b[i,3]<0.05)

}

上述计算的想法是重复5000次随机分组，每次随机分组以后，对基线疾病严重程度与分得的AB两组之间做卡方检验和Fisher确切概率法检验。检验后看这重复进行的5000次随机分组后，组间检验p<0.05的次数所占的比例有多少。计算结果如下：

	p≥0.05	p<0.05	p<0.05构成比
卡方检验	4724	276	5.52%
确切概率法	4733	267	5.34%

上述模拟结果显示，随机分组后组间基线疾病严重程度差异有统计学意义的发生概率为5.34-5.52%。在概率统计上这是一件接近小概率事件的事件。

那么如果采用分层随机分组会是什么结果呢?我们将研究对象分为轻度、中度和重度3层，每1层内分别做完全随机分组，然后再将数据合并起来分析，如此重复5000次。

其结果如下：

	p≥0.05	p<0.05	p<0.05构成比
卡方检验	5000	0	0.00%
确切概率法	5000	0	0.00%

结果差异有统计学意义的随机分组为0次，为什么?因为咱们把每1层的人数都等数量地随机分到两组中去，所有计算的卡方值永远为0，p值则为1。从上述的模拟可以看出分层等数量随机分组后，分层因素在组间差异一定没有统计学意义，而完全随机分组后基线资料有统计学差异接近小概率事件。

2、不同样本量情况下是否有影响

下面我们再模拟一下不同样本量的情况下是否会对结果有影响，此次模拟1000次。代码如下：

numb2=c(60,120,180,240,300)

dat=data.frame(rep(-1,1000))

dat2=data.frame(rep(-1,1000))

```
dat3=data.frame(rep(-1,1000))

dat4=data.frame(rep(-1,1000))

for(numb in numb2){

grade=c(rep(1, numb/2), rep(2, numb/3), rep(3, numb/6))

set.seed(20180101)

bg=rep(c(0,1), numb/2)

b=data.frame(rep(-1,1000))

for(i in 1:1000){

set.seed(i)

group=sample(bg, numb,rep=F)

a1= chisq.test(grade, group)

a2= fisher.test(grade,group)

b[i,1]=a1$p.value

b[i,2]=(b[i,1]<0.05)

b[i,3]=a2$p.value

b[i,4]=(b[i,3]<0.05)

}

dat[,numb/60]=b[,1]

dat2[,numb/60]=b[,2]

dat3[,numb/60]=b[,3]

dat4[,numb/60]=b[,4]

}
```

上述计算的将研究的样本量设置为60-300不同的5个类型。然后在每一个样本量下重复1000次随机分组，每次随机分组以后，对基线疾病严重程度与分得的AB两组之间做卡方检验和Fisher确切概

率法检验。检验后看这重复进行的1000次随机分组后，组间检验p<0.05的次数所占的比例有多少。计算结果如下：

	样本量	p≥0.05	p<0.05	p<0.05构成比
卡方检验	60	959	41	4.10%
	120	947	53	5.30%
	240	951	49	4.90%
	480	953	47	4.70%
	600	937	63	6.30%
确切概率法	60	951	49	4.90%
	120	958	42	4.20%
	240	951	49	4.90%
	480	955	45	4.50%
	600	941	59	5.90%

这个结果是不是似曾相识，是的，差异有统计学意义的发生概率又接近5%。

那么为什么会这样呢?因为这是完全随机分组，分组后你再对其进行检验p值，理论上p的取值应该是个均匀分布。然后你期望去计算p<0.05的比例是多少?那肯定是接近0.05啊!

3、小结

通过上面的模拟，再试想一下一开始的研究设计的问题：不同严重程度在随机分组后出现组间差异有统计学意义的概率有多大?这就是一个小概率事件。如果采用分层随机分组，被分层的因素在组间还会有差异吗?不会了。但是仅仅是被分层的因素没有差异，没有被分层的因素出现差异的概率依然是小概率事件。