

基于 Wald 检验、逻辑回归、多重填补、条件把握度、 α 消耗函数、Graphical Approach 和广义估计方程等多种统计方法的三期临床试验数据分析报告

目录

1 问题一	1
1.1 问题 1.1	1
1.2 问题 1.2	2
1.3 问题 1.3	3
1.4 问题 1.4	3
2 问题二	7
2.1 问题 2.1	7
2.2 问题 2.2	8
2.2.1 问题 2.2.1	8
2.2.2 问题 2.2.2	9
2.3 问题 2.3	9
3 问题三	10
3.1 问题 3.1	10
3.2 问题 3.2	10
3.3 问题 3.3	11
4 问题四	11

1 问题一

1.1 问题 1.1

根据 ICH E9(R1) 的有关定义，我们提出的估计目标为：在满足入排标准的人群中，较安慰剂组而言，该药物在第 24 周时通过患者评分标准 A 判定的疾病较基线的改善情况。

该估计目标的五个属性如下：

1. 治疗：受试者从第 0 周开始每隔 4 周到研究中心进行给药并接受定期检查。
2. 人群：满足入排标准，经过分层随机分配到药物组和安慰剂组的所有受试者。

3. 变量：受试者在第 24 周时通过患者评分标准 A 判定的疾病较基线是否改善（二分类变量）。
4. 伴发事件：1) 患者在 24 周前终止治疗并退出试验；2) 患者在 24 周之内采取了额外治疗，如背景治疗改变、补救治疗等。
5. 群体层面汇总：统计药物组和安慰剂组在第 24 周时经患者评分标准 A 判定为疾病较基线有所改善的受试者比例（以下简称应答率），并求出两组间应答率差值的 95% 可信区间。

该估计目标较为合理，因为该目标能精确描述为实现研究目标而需要估计的治疗效应，并在群体水平上汇总比较同质性患者在不同治疗条件下的结局，能够反映出该药物相比于安慰剂在治疗有效性和安全性上的差异。

1.2 问题 1.2

根据问题 1.1 所述估计目标，我们认为这是一个对二分类资料的分析问题。

首先，应采用卡方检验，根据两组受试者中应答者和非应答者数量列出四格表如下：

	无应答	应答	合计
安慰剂组	a	b	a+b
药物组	c	d	c+d
合计	a+c	b+d	a+b+c+d

通过计算对应的 p 值，可以检测药物组和安慰剂组的应答率差异是否具有统计学意义。

其次，可以借助 **Wald test**，估计两组应答率的差异。对于总体应答率 π 而言，当样本量 n 足够大，样本应答率 P 不太小， $(1 - P)$ 也不会太小，比如 $n(1 - P)P > 5$ 时，根据中心极限定理， P 的抽样分布可以用正态分布近似，也即：

$$P \rightarrow \mathcal{N}\left(\pi, \frac{P(1 - P)}{n}\right)$$

记 P_0, P_1 分别为安慰剂组和药物组的样本应答率，那么

$$S_0 = \sqrt{\frac{P_0(1 - P_0)}{n_0}}, S_1 = \sqrt{\frac{P_1(1 - P_1)}{n_1}}$$

分别为安慰剂组与药物组的样本应答率的标准误。根据概率论的知识，当两组样本量比较大时，两组的比例差异近似于服从下述正态分布：

$$P_1 - P_0 \rightarrow \mathcal{N}(\pi_1 - \pi_0, S_0^2 + S_1^2)$$

进一步可以使用单侧的 Wald test，检测安慰剂组和药物组的应答率的差异是否有统计学意义，并估计两组应答率差值的 95% 置信区间为：

$$P_1 - P_0 \pm Z_{\alpha/2} \sqrt{S_0^2 + S_1^2}$$

再者，也可以考虑建立多元逻辑回归模型。以分组（药物组、安慰剂组）作为自变量，考虑到分层因素，可以再加入所在中心（亚洲、欧洲、北美洲）和体重（轻重）作为协变量，建

立多元逻辑回归模型如下：

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

其中：

X_1 ：表示体重，0 表示对应基线体重为轻，1 表示对应基线体重为重。

X_2, X_3 ：表示地区， $X_2 = 0, X_3 = 0$ 表示亚洲， $X_2 = 1, X_3 = 0$ 表示欧洲， $X_2 = 0, X_3 = 1$ 表示北美洲。

X_4 ：表示分组，0 表示安慰剂组，1 表示药物组。

然后再通过逐步回归的方法，去除影响不显著的变量。该模型能够排除地区和体重所带来的干扰。此外， $\exp(\beta_4)$ 即为对应于受试者分组的的优势比 OR_4 。当该优势比大于 1 的时候，可以认为药物组相较于安慰剂组确实更有效。由于 β_4 是未知参数，可以基于观察数据结合模型求出 β_4 的估计值 b_4 和标准误 $Se(b_4)$ ，得到对应于分组情况 X_4 的总体优势比 OR_4 的 $1 - \alpha$ 可信区间

$$\exp[b_j \pm Z_{\alpha/2} Se(b_j)]$$

1.3 问题 1.3

我们假定数据缺失机制为 MAR。我们认为，可以使用多重逻辑回归填补法对缺失数据进行填补。先用所有在第 24 周观测到的数据，得到问题 1.2 中提出的逻辑回归模型。将每个缺失第 24 周数据的受试者体重、地区和所在组别带入模型，求出其第 24 周应答的概率 P ，再生成多个应答率为 P 的服从伯努利分布随机数，作为该缺失者的多个填补值，从而得到多个填补后的完整数据集。然后，使用鲁宾法则，将各数据集的结果汇总，得到最终的两组应答率之差的估计值、标准误、置信区间，以及用填补数据拟合出的多元逻辑回归模型。将这些填补后所得结果与未考虑缺失数据所得结果进行比较，若二者接近，则可以证明主要分析结果的稳健性。以上填补过程可以使用 R 语言中的 mice 包完成。

我们认为，这种基于多重逻辑回归的后验填补方法，可以保证结果具有较小的偏差和较大的有效性。

至于临界点方法，应往“对原分析结果更不利”的方向考虑，假设药物组和安慰剂组中退出试验的受试者的应答率，然后计算在该应答率下的假设检验结果，在本题中，即为卡方检验是否有显著性差异，以及两组样本应答率之差的置信区间是否包含 0。我们分别调整两组中提前退出的受试者的应答率，找到使得上述检验结果发生反转的临界值，这个临界值就是一个临界点。由于我们允许两组中假设的提前退出者的应答率可以不同，因此应该有不只一个临界点，我们推测临界点应分布于某条曲线附近。

1.4 问题 1.4

首先，我们对未缺失部分的数据进行卡方检验，列出四格表如下：

	无应答	应答	合计
药物组	33	52	85
安慰剂组	55	30	85
合计	88	82	170

利用 R 语言进行 Yates' 连续性校正后的卡方检验，结果为

$$\chi^2 = 10.389, p = 0.001267 < 0.05$$

按 $\alpha = 0.05$ 的水准拒绝原假设，差别有统计学意义，可以认为药物组应答率高于安慰剂组。

下面，我们近似估计两组应答率之差的置信区间。由上述四格表可知：

$$\begin{aligned} P_0 &= \frac{30}{85} = 0.3529 \\ P_1 &= \frac{52}{85} = 0.6118 \\ P_1 - P_0 &= 0.6118 - 0.3529 = 0.2589 \\ S_0 &= \sqrt{\frac{P_0(1 - P_0)}{n_0}} = 0.05183 \\ S_1 &= \sqrt{\frac{P_1(1 - P_1)}{n_1}} = 0.05286 \\ S &= \sqrt{S_0^2 + S_1^2} \\ &= \sqrt{0.05183^2 + 0.05286^2} \\ &= 0.07403 \\ CI_{0.95} &= P_1 - P_0 \pm Z_{\alpha/2} \sqrt{S_0^2 + S_1^2} \\ &= (0.1137, 0.4039) \end{aligned}$$

注意到 95% 置信区间不包含 0，这等价于按 $\alpha = 0.05$ 的水准拒绝原假设，可以认为药物组应答率高于安慰剂组。

最后，我们对未缺失的数据进行逻辑回归分析，并进行逐步回归，去除影响不显著的变量，R 代码以及回归结果如下：

Listing 1: 建立逻辑回归模型并展示回归结果

```
1 mydata = data[which(data$week=='week 24'&data$y1!='NA'),]
2 mydata$`Assigned Treatment` = factor(mydata$`Assigned Treatment`,levels=c('treatPlacebo','
3 Experimental Drug'))
4 myglm0 = glm('y1~weight+region+treat',family=binomial(link = 'logit'),data=mydata)
5 myglm = step(myglm0,trace=F) #逐步回归
6 summary(myglm) # 查看回归结果
7 # Call:
8 # glm(formula = y1 ~ region + treat, family = binomial(), data = mydata)
9 #
10 # Deviance Residuals:
11 #   Min       1Q   Median       3Q      Max
12 # -1.5903  -1.1305  -0.8182   1.1228   1.5856
13 #
14 # Coefficients:
15 #              Estimate Std. Error z value Pr(>|z|)
16 # (Intercept)      0.15428    0.34152   0.452  0.65146
17 # regionEurope      0.77835    0.39575   1.967  0.04921 *
18 # regionNorth America -0.03275    0.41373  -0.079  0.93690
19 # treatPlacebo     -1.04392    0.32400  -3.222  0.00127 **
```

```

20 # ---
21 # Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
22 #
23 # (Dispersion parameter for binomial family taken to be 1)
24 #
25 # Null deviance: 235.46  on 169  degrees of freedom
26 # Residual deviance: 218.07  on 166  degrees of freedom
27 # AIC: 226.07
28 #
29 # Number of Fisher Scoring iterations: 4

```

注意到在逐步回归后，体重变量被排除在外，说明体重对应答率的影响并不显著。我们可以将逻辑回归模型写成如下具体形式：

$$\log\left(\frac{p_i}{1-p_i}\right) = 0.15428 + 0.77835X_2 + -0.03275X_3 + 1.04392X_4$$

其中：

X_2, X_3 ：表示地区， $X_2 = 0, X_3 = 0$ 表示亚洲， $X_2 = 1, X_3 = 0$ 表示欧洲， $X_2 = 0, X_3 = 1$ 表示北美洲。

X_4 ：表示分组，0 表示安慰剂组，1 表示药物组。

进一步，可以计算优势比 OR_4 的 $1 - \alpha$ 的可信区间。取 $\alpha = 0.05$ ，得到

$$\exp[b_j \pm Z_{\alpha/2}Se(b_j)] = (1.505, 5.360)$$

下面，进行多重填补和临界点分析。

先调用 mice 包，按 1.3 中所述方法基于多重逻辑回归进行多重填补，并利用鲁宾法则合并各填补后数据集，得到合并后的应答率差值以及置信区间。R 代码以及结果如下：

Listing 2: 基于多元逻辑回归进行多重填补并用鲁宾法则合并结果

```

1 dataf1 = data[which(data$week=='week 24'),3:6]
2 colnames(dataf1) = c('weight','region','treat','y1')
3 dataf1$y1 = as.logical(dataf1$y1)
4 dataf1 = as.data.frame(dataf1)
5 imp_data = mice(dataf1,method='logreg') # 按多元逻辑回归的方式进行插补
6 fit = with(imp_data,glm(y1~ region+treat,family = binomial(link='logit')) # 用插补后的数据拟合多元逻辑回归模型
7 summary(pool(fit)) # 查看回归结果
8
9 # 用插补后依据Rubin法则合并后的数据
10 # 以地区和分组作为自变量，进行多元逻辑回归的结果如下：
11 # term estimate std.error statistic    df p.value
12 # 1      (Intercept)  0.14576    0.3319    0.4392  75.17 0.661766
13 # 2      regionEurope  0.75691    0.3931    1.9254  76.82 0.057882
14 # 3 regionNorth America -0.03036    0.4092   -0.0742 101.74 0.940996
15 # 4      treatPlacebo -1.00828    0.3292   -3.0627  69.30 0.003122
16 # 可以发现，该结果与用完整数据进行拟合的效果基本一致
17

```

```

18 m0 = c(0,0,0,0,0)
19 m1 = m0
20 v = m0
21 for(i in 1:5){
22   cmplt = complete(imp_data,i)
23   p0 = mean(cmplt[which(cmplt$treat=='Placebo'),]$y1)
24   p1 = mean(cmplt[which(cmplt$treat!='Placebo'),]$y1)
25   m0[i] = p0
26   m1[i] = p1
27   v[i] = p0*(1-p0)/100 + p1*(1-p1)/100 # 应答率之差的variance
28 }
29 P = pool.scalar(m1-m0,v,n=200,k=2) # 合并多个插补数据集的应答率之差以及标准误
30 dff = P$qbar
31 # dff = 0.236
32 S = sqrt(P$ubar)
33 # S = 0.06861
34 CI_95 = c(dff-1.96*S,dff+1.96*S)
35 # CI_95 = (0.1015 0.3705) 95%置信区间不包含0, 仍拒绝原假设

```

可以发现，将填补后的数据合并后进行多元逻辑回归，结果基本一致。特别的，在分组变量上系数为-1.00828，与原数据回归结果所得系数-1.04392 基本一致，可以说明用原数据所得的回归结果具有稳健性。

此外，用填补后的 5 组数据合并后所得的两组应答率之差的点估计为 $\tilde{P}_1 - \tilde{P}_0 = 0.236$ ，95% 的置信区间为 (0.1015, 0.3705)，与用原数据所得的估计结果也较为接近，并且同样能拒绝原假设，说明两组的应答率差异具有统计学意义。这也证明了用原数据进行假设检验的结果具有稳健性。

再进行临界点分析。分别假设安慰剂组提前退出试验的受试者中，有 0 到 15 人属于应答者（该数量记作 m_0 ）则有 $15-m_0$ 人是非应答者；假设药物组提前退出试验的受试者中，有 0 到 15 人属于非应答者（该数量记作 m_1 ），则有 $15-m_1$ 是应答者。注意到，当 m_0 越大，或者 m_1 越大的时候，检验的结果将会对原分析结果“更为不利”。示意表格如下：

	无应答	应答	合计
药物组	$33+m_1$	$52+(15-m_1)$	100
安慰剂组	$55+(15-m_0)$	$30+m_0$	100
合计	$103+m_1-m_0$	$97-m_1+m_0$	200

用 R 语言对上述 256 种填补情况进行卡方检验，并考察两个总体应答率之差的 95% 置信区间是否包含 0（R 代码详见附录），得到临界点分析示意图如下：

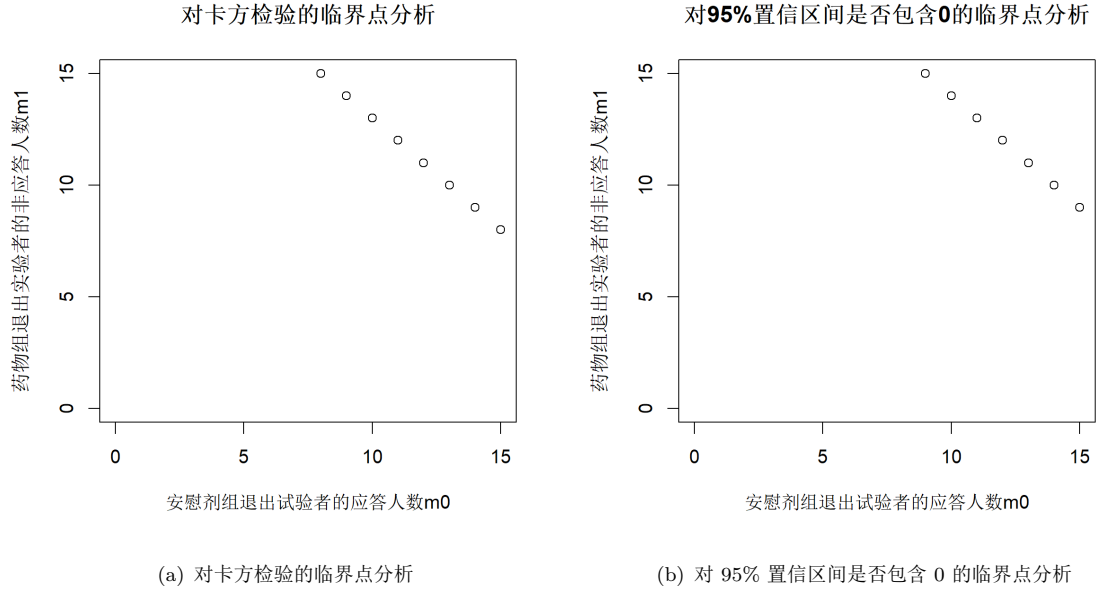


图 1: 临界点分析

图中点即为对应检验的结果发生反转的临界点。可以看出基本分布在一条直线上。在这条直线左下方的范围内, 检测结果不变。也即在大部分取值情况下, 检验结果与原分析结果是一致的, 这说明了原分析结果具有稳健性, 也说明缺失数据对分析结果的影响较小。

2 问题二

2.1 问题 2.1

在以下讨论中, 我们仅考虑平衡设计, 即两组样本量相等情况。此外, 我们认为对于三个终点的讨论, 都建立在脱落率 (即中途停止治疗的受试者比例) 都为 15% 的假设下。

样本量估计 原假设和备择假设分别为:

$$H_0: \pi_1 - \pi_0 = 0$$

$$H_1: \pi_1 - \pi_0 > 0$$

设单侧检验一类错误概率为 $\alpha = 0.025$, 二类错误概率为 $\beta = 0.05$, 药物组 $\pi_1 = 0.55$, 安慰剂组 $\pi_0 = 0.3$, 在不考虑脱落的情况下, 可以得到总样本量 (两组样本量之和) 为:

$$\begin{aligned} n &= 2 \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{(\pi_1 - \pi_0)^2} [\pi_1(1 - \pi_1) + \pi_0(1 - \pi_0)] \\ &= 2 \frac{(1.96 + 1.64)^2}{(0.55 - 0.30)^2} [(0.55(1 - 0.55)) + 0.30(1 - 0.30)] \\ &= 189.8 \end{aligned}$$

向上取整得到 $n = 190$

设脱落率为 $\eta = 0.15$, 则考虑脱落率后的样本量应该为:

$$n^* = \frac{n}{1 - \eta} = \frac{190}{1 - 0.15} = 223.5$$

向上取整得到 $n^* = 224$.

计算次要终点的把握度 我们仍假设脱落率为 $\eta = 0.15$ ，则应取考虑脱落后的总样本量 $n = 190$ 进行计算。对于第一个次要终点，原假设和备择假设分别为：

$$H_0 : \mu_1 - \mu_0 = 0$$

$$H_1 : \mu_1 - \mu_0 > 0$$

设 $\mu_1 = 0.65$, $\mu_2 = 0.35$ ，则两样本应答率之差的标准误为

$$\begin{aligned} S &= \sqrt{\frac{\mu_1(1-\mu_1)}{n/2} + \frac{\mu_0(1-\mu_0)}{n/2}} \\ &= \sqrt{\frac{0.65(1-0.65)}{95} + \frac{0.35(1-0.35)}{95}} \\ &= 0.0692 \end{aligned}$$

统计量的单侧界值为

$$c = Z_{1-\alpha}S = 1.96 \times 0.0692 = 0.1356$$

得到把握度为

$$\begin{aligned} 1 - \beta &= \Phi \left[\frac{\mu_1 - \mu_0 - c}{S} \right] \\ &= \Phi \left[\frac{0.65 - 0.35 - 0.1356}{0.0692} \right] \\ &= 99.1\% \end{aligned}$$

其中 $\Phi(\cdot)$ 是标准正态分布的累积分布函数。

类似的，对于第二个次要终点，设 $\lambda_1 = 0.74$, $\lambda_2 = 0.50$ ，则两样本应答率之差的标准误为

$$\begin{aligned} S &= \sqrt{\frac{\lambda_1(1-\lambda_1)}{n/2} + \frac{\lambda_0(1-\lambda_0)}{n/2}} \\ &= \sqrt{\frac{0.74(1-0.74)}{95} + \frac{0.50(1-0.50)}{95}} \\ &= 0.0682 \end{aligned}$$

统计量的单侧界值为

$$c = Z_{1-\alpha}S = 1.96 \times 0.0682 = 0.13375$$

得到把握度为

$$\begin{aligned} 1 - \beta &= \Phi \left[\frac{\lambda_1 - \lambda_0 - c}{S} \right] \\ &= \Phi \left[\frac{0.74 - 0.50 - 0.13375}{0.0682} \right] \\ &= 94.0\% \end{aligned}$$

2.2 问题 2.2

2.2.1 问题 2.2.1

根据陈建平等的研究 [2]，我们决定基于布朗运动理论，估计条件把握度，再通过二分搜索法估计后一阶段所需样本含量。

该算法可以根据期中数据，计算条件把握度，如果把握度太低，则调整样本含量以达到期望把握度，如果条件把握度较高，则建议提前结束临床试验。由于期中分析涉及到多重性问题，需要对一类错误进行控制。Lan 和 DeMets 提出了五类 α 消耗函数 [1]，可以根据具体情况选用其中一种。在本算法中，我们采用了较为保守但总期望样本量较低的 O'Brien-Fleming 消耗函数

$$\alpha(t) = 1 - \Phi\left(\frac{Z_{1-\alpha}}{\sqrt{t}}\right)$$

对一类错误进行控制，其中 t 是信息时间。

该算法输入为：试验前预计总样本量 N ，已完成的样本量 n ，试验允许的最大样本量 N_{\max} ，两组各自的应答率 η_0, η_1 ，总的单侧检验显著性 α （默认为 0.025），把握度 $1 - \beta$ （默认为 0.95），迭代中止标准 tol （即条件把握度与期望把握度的最大容忍误差）。该算法输出包括如下四种情况：

1. 重估样本量大于试验允许最大样本量 N_{\max} ，结束试验，无效终止，试验失败。
2. 给出介于 n 和 N_{\max} 的重估样本量 N' 。
3. tol 设置过小，无法满足二分法迭代停止准则，需重新设置。

设计算法时采用的统计原理以及具体算法步骤请参考附件 1，完成该算法的 Python 代码请参考附件 2。

2.2.2 问题 2.2.2

如果在实际临床试验中通过样本量重估的样本量小于实际预设定的样本量，应按照下述步骤处理：

- 首先，检查样本量重估的过程是否有误，包括数据清洗过程中是否出现漏误，算法和代码是否合理，是否考虑了分层随机化与多中心因素，以及是否考虑排除数据中的离群值。
- 其次，如果以上诸检查均无误，应该向独立数据监察委员会汇报结果。
- 然后，根据陈峰和夏结来等人的专著 [1]，当新样本量小于等于已完成的样本量 n 时，则可以提前终止试验；但当新样本含量小于原计划的样本量但大于已完成的样本量 n ，为确保试验的把握度，我们仍然会建议采用方案原计划的样本量。
- 若 IDMC 决定采用新样本量，则在最终分析时，应该在统计报告中详细说明样本量重估的过程、结果与可能影响。

2.3 问题 2.3

此处仅考虑对两组应答率之差的单侧 Wald 检验方法。设信息时间为 $\tau' = \frac{n}{N'}$ ，其中 N' 是重估后的样本量。根据陈建平等人的研究 [2]，在进行最终分析时，需要考虑期中分析带来的一类错误膨胀问题。这里我们与上文一致，仍采用五种类型消耗函数中的 O'Brien-Fleming 消耗函数 $\alpha(\tau) = 1 - \Phi\left(\frac{Z_{1-\alpha}}{\sqrt{t}}\right)$ （当然，也可以替换为其他消耗函数，只需要与期中分析时采用的类型保持一致即可）。最终分析的显著性水平应该设置为 $\alpha_2 = 0.025 - \alpha(\tau')$ 。

那么，最终分析的单侧界值则相应变为 $c' = Z_{1-\alpha_2}$ ，也即当

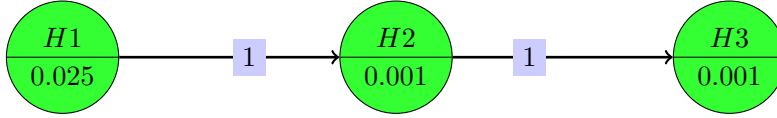
$$\frac{P_1 - P_0}{S} > Z_{1-\alpha_2}$$

时，可以拒绝原假设，认为药物组应答率高于安慰剂组，并且仍然能把一类错误率控制在单侧 0.025 的水平上。

3 问题三

3.1 问题 3.1

我们决定采用回退法 (fallback) 进行多重性假设检验，考虑 α_i 的回收与分配后，以 graphical approach 方法展示假设检验的流程如下：



在该检验中，在初始时分配 $\alpha_1 = 0.023, \alpha_2 = 0.001, \alpha_3 = 0.001$ ，并在拒绝上一个假设后，将所有的 α 值都传递给下一个假设。

我们提出该多重性假设检验方法的合理性如下：

1. 主要终点假设检验重要性最高，故将其作为第一个假设进行检验。
2. 由于可以进行 α 的回收与分配，故使用贪心算法的思想，在初始时即让 α_1 达到在 $\alpha_i > 0.001$ 约束下的最大值，也即 $\alpha_1 = 0.023$ 。当主要终点的原假设被拒绝后，这部分显著性水平可以被全部传递到下一个假设，使下一个假设也更容易被拒绝。
3. 初始时三个假设所分配的显著性水平相加正好为 0.025，故该检验流程能将总的一类错误控制在单侧 0.025。
4. 每个终点的原假设被拒绝后，都将该终点的所有显著性水平传递给下一个终点，这样能够尽可能地穷耗总的 α 。
5. 根据问题 2.1 里的计算结果，第一个次要终点的把握度高于第二个次要终点，也即第一个次要终点的原假设更可能被拒绝。因此先检验第一个次要终点，再检验第二个次要终点，更有可能让两个次要终点的原假设都被拒绝，增加该药获批上市的可能性。

3.2 问题 3.2

在问题 3.3 中，根据我们提出的 graphical approach 方法，将显著性水平设置为单侧 $\alpha_1 = 0.023$ 。同样使用问题 2.1 中的样本量计算公式，可以得到：

$$\begin{aligned}
 n &= 2 \frac{(Z_{1-0.023} + Z_{1-0.05})^2}{(\pi_1 - \pi_0)^2} [\pi_1(1 - \pi_1) + \pi_0(1 - \pi_0)] \\
 &= 2 \frac{(1.995 + 1.64)^2}{(0.55 - 0.30)^2} [(0.55(1 - 0.55)) + 0.30(1 - 0.30)] \\
 &= 193.4
 \end{aligned}$$

向上取整得到 $n = 194$

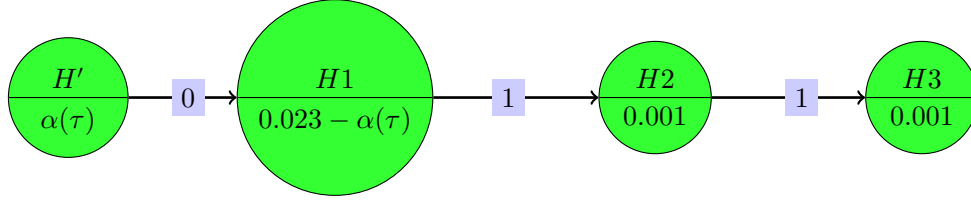
设脱落率为 $\eta = 0.15$ ，则考虑脱落率后的样本量应该为：

$$n^* = \frac{n}{1 - \eta} = \frac{194}{1 - 0.15} = 228.2$$

向上取整得到 $n^* = 229$ 。

3.3 问题 3.3

在进行其中分析后, graphical approach 的思想仍然是适用的, 但需要对上述的多重假设检验流程进行修改, 在主要终点的检验之前再增加一个代表期中分析的节点, 并同样利用 α 消耗函数, 将总一类错误概率分配一部分给期中分析。此外, 在期中分析之后, 不会把期中分析 α 传递给最终检验。示意图如下:



其中, H' 代表了期中分析, τ 是信息时间, $\alpha(\tau)$ 是某个预先选取的 α 消耗函数, 并且需要满足 $0.023 - \alpha(\tau) > 0.001$ 。

4 问题四

对于重复测量数据, 不同时间点的测量数据之间存在相关性, 且中途部分受试者退出, 造成部分数据缺失, 这使得我们无法使用传统的方差分析。我们认为, 对于上述的问题, 可以使用广义估计方程 (GEE) 进行模型拟合, 该模型有较稳健且一致的估计。此外, 我们推测组别和测量时间之间很可能具有交互效应, 也即药物组和安慰剂组的应答率随时间的增加具有不同的变化趋势。因此, 还可以在 GEE 中引入组别和时间的交互项。下面, 我们仍以主要终点为例, 对附件中的试验数据建模, 有关的 R 代码与建模结果如下:

Listing 3: 以主要终点为例, 对重复测量数据用 GEE 进行建模

```

1 library(gee)
2 myData = data[,1:6]
3 myData$week = factor(myData$week,levels=c('week 4','week 8','week 12','week 16','week 20','
      week 24'))
4 colnames(myData) = c('id','week','weight','region','treat','y1')
5 fit <- gee(y1~week+weight+region+treat+treat*week,id=id,data=myData,corstr = 'unstructured',
      family = 'binomial')
6 summary(fit)
7
8 # 利用广义估计方程拟合模型的主要结果如下
9 # Model:
10 #   Link:                               Logit
11 # Variance to Mean Relation: Binomial
12 # Correlation Structure:      Unstructured
13 #
14 # Call:
15 #   gee(formula = y1 ~ week + weight + region + treat + treat * week,
16 #       id = id, data = myData, family = "binomial", corstr = "unstructured")
17 #
18 # Summary of Residuals:
19 #   Min       1Q   Median       3Q      Max
20 # -0.7104984 -0.3027152 -0.1552428  0.4124540  0.9236200
21 #
22 # Coefficients:

```

```

23 # Estimate Naive S.E. Naive z Robust S.E. Robust z
24 # (Intercept) -1.86199429 0.3496230 -5.3257196 0.3648092 -5.1040228
25 # weekweek 8 0.88836188 0.2311107 3.8438806 0.2457702 3.6146041
26 # weekweek 12 1.46818995 0.2823656 5.1996056 0.2771771 5.2969390
27 # weekweek 16 1.72488443 0.2980612 5.7870136 0.2880531 5.9880775
28 # weekweek 20 1.91942309 0.3198643 6.0007423 0.3098292 6.1951000
29 # weekweek 24 2.01216619 0.3376028 5.9601580 0.3260655 6.1710493
30 # weightWeight group 2 -0.23614495 0.2506380 -0.9421755 0.2506232 -0.9422309
31 # regionEurope 0.74763409 0.2993695 2.4973620 0.2964011 2.5223732
32 # regionNorth America 0.05580632 0.3234823 0.1725174 0.3321409 0.1680200
33 # treatPlacebo -0.39443994 0.4168942 -0.9461392 0.4149862 -0.9504893
34 # weekweek 8:treatPlacebo -0.59953101 0.3598567 -1.6660273 0.3628223 -1.6524093
35 # weekweek 12:treatPlacebo -0.72547594 0.4327674 -1.6763646 0.4345050 -1.6696608
36 # weekweek 16:treatPlacebo -0.78319055 0.4526123 -1.7303785 0.4454338 -1.7582647
37 # weekweek 20:treatPlacebo -1.00887956 0.4854820 -2.0780988 0.4725623 -2.1349135
38 # weekweek 24:treatPlacebo -0.57184144 0.5011783 -1.1409940 0.4927536 -1.1605017
39 #
40 # Estimated Scale Parameter: 1.032417

```

注意到该模型的地区变量的系数（Europe: 0.74763, North America: 0.05581）和问题 1.4 中只用第 24 周末缺失数据进行多元逻辑回归（逐步回归）所得到的地区变量系数（Europe: 0.77835, North America: -0.03275）整体上一致，这反映出了 GEE 模型的稳健性和一致性。

但是，在引入时间和分组的交叉项之后，分组变量的系数变为 0.39443994，与问题 1.4 中所得到的分组变量系数 1.04392 有较大差异，这反映出时间与分组之间存在比较明显的交互作用，可以看作原本的分组变量引起的效应，被进一步拆解为独立效应和交互效用。这也说明了两组的应答率随时间的变化趋势不同。

此外，如果将该模型的分组变量，以及分组与第 24 周的交互项的系数相加，即 $0.57184144 + 0.39443994 = 0.96628138$ ，发现该值与问题 1.4 中得到的分组变量系数 1.04392 较为接近，说明在使用 GEE 进行分析后，将分组变量的独立效应和与时间的交互效应相加，基本与不考虑时间因素所得到的分组变量效应相一致，这同样说明了 GEE 模型的稳健性和一致性。

参考文献

- [1] 陈峰 and 夏结来. 临床试验统计学. 临床试验统计学, 2018.
- [2] 陈建平, 魏永越, 陈峰, and 于浩. 期中分析的条件把握度及样本含量再估计. 中国卫生统计, (4):3, 2010.

附录：正文与附录中均未包含的代码

临界点分析所使用的 R 代码

Listing 4: 临界点分析所使用的 R 代码

```

1 chi.TF = function(a,b,c,d){
2   t2 = rbind(c(a,b),c(c,d))
3   # print(t2)
4   xt = chisq.test(t2);

```

```

5   # print(xt$p.value)
6   xt$p.value < 0.05
7 }
8
9 CI.TF = function(a,b,c,d){
10   p0 = d/(c+d);
11   p1 = b/(a+b);
12   s0 = sqrt(p0*(1-p0)/(c+d));
13   s1 = sqrt(p1*(1-p1)/(a+b));
14   CI.left = p1-p0-1.96*sqrt(s0^2 + s1^2);
15   0 < CI.left
16 }
17
18
19 a = 33; b = 52; c = 55; d = 30;
20
21 m0.list = c()
22 m1.list = c()
23 for (m0 in 0:15){
24   d.new = d+m0
25   c.new = c + 15 - m0
26   for (m1 in 0:15){
27     a.new = a + m1
28     b.new = b + 15 - m1
29     # print(c(m0,m1))
30     if (m1==0){
31       state = chi.TF(a.new,b.new,c.new,d.new)
32       next
33     }
34     if (chi.TF(a.new,b.new,c.new,d.new)!=state){
35       m0.list = append(m0.list,m0)
36       m1.list = append(m1.list,m1)
37       state = chi.TF(a.new,b.new,c.new,d.new)}
38   }}
39
40 plot(m0.list,m1.list,main = '对卡方检验的临界点分析',
41      xlab = '安慰剂组退出试验者的应答人数m0',
42      ylab = '药物组退出实验者的非应答人数m1',
43      xlim = c(0,15),ylim=c(0,15))
44
45
46
47 m0.list = c()
48 m1.list = c()
49 for (m0 in 0:15){
50   d.new = d+m0
51   c.new = c + 15 - m0
52   for (m1 in 0:15){
53     a.new = a + m1
54     b.new = b + 15 - m1
55     # print(c(m0,m1))
56     if (m1==0){

```

```

57     state = CI.TF(a.new,b.new,c.new,d.new)
58     next
59 }
60 if (CI.TF(a.new,b.new,c.new,d.new)!=state){
61     m0.list = append(m0.list,m0)
62     m1.list = append(m1.list,m1)
63     state = CI.TF(a.new,b.new,c.new,d.new)}
64 }}
65
66 plot(m0.list,m1.list,main = '对95%置信区间是否包含0的临界点分析',
67      xlab = '安慰剂组退出试验者的应答人数m0',
68      ylab = '药物组退出实验者的非应答人数m1',
69      xlim = c(0,15),ylim=c(0,15))

```