

器械中有大一类产品是诊断产品，譬如，试剂盒、超声、核磁共振、X 射线等。诊断产品做临床试验主要验证的是诊断准确率，譬如，灵敏度、特异度、阳性预测率、阴性预测率、ROC 曲线下面积等指标。如果一个新的产品要通过做临床试验提交药监局批准上市，需要考虑常见的几种诊断试验设计类型。

根据有无金标准，设计类型可以分为以下四种类型：

- (1) 有金标准的单组设计：
- 此时主要衡量新诊断产品准确率是否达到一定标准。譬如，通过做临床试验验证乳腺超声的灵敏度是否大于 0.85，特异度是否大于 0.8。

		金标准		
新诊断产品		有病	无病	
	阳性	a	b	$a + b$
	阴性	c	d	$c + d$
		$a + c$	$b + d$	$a + b + c + d$

灵敏度= $\frac{a}{a+c}$

特异度= $\frac{d}{b+d}$

可以采用精确概率法或者近似正态 Z 检验对灵敏度和特异度是否达到目标值进行检验。

- (2) 有金标准的两组配对设计：
- 此时，主要衡量新诊断产品是否优于或者非劣于对照产品。譬如，通过做临床试验验证 3D 乳腺 X 射线的灵敏度和特异度是否优于 2D 乳腺射线。

已经过金标准确诊为疾病的受试者				
		对照产品		
新诊断产品		阳性	阴性	
	阳性	a_1	b_1	$a_1 + b_1$
	阴性	c_1	d_1	$c_1 + d_1$
		$a_1 + c_1$	$b_1 + d_1$	$a_1 + b_1 + c_1 + d_1$

新诊断产品的灵敏度= $\frac{a_1+b_1}{a_1+b_1+c_1+d_1}$

对照产品的灵敏度= $\frac{a_1+c_1}{a_1+b_1+c_1+d_1}$

若要比 较新诊断产品的灵敏度是否优于对照产品，可以采用 McNemar 检验（想一想为什么？）

$$\chi^2_{McNemar} = \frac{(b_1 - c_1)^2}{b_1 + c_1}$$

已经过金标准确诊为健康的受试者				
		对照产品		
新诊断产品		阳性	阴性	
	阳性	a_2	b_2	$a_2 + b_2$
	阴性	c_2	d_2	$c_2 + d_2$
		$a_2 + c_2$	$b_2 + d_2$	$a_2 + b_2 + c_2 + d_2$

新诊断产品的特异度= $\frac{c_2+d_2}{a_2+b_2+c_2+d_2}$

对照产品的特异度= $\frac{b_2+d_2}{a_2+b_2+c_2+d_2}$

若要比 较新诊断的特异度是否优于对照产品，可以采用 McNemar 检验：

$$\chi^2_{McNemar} = \frac{(b_2 - c_2)^2}{b_2 + c_2}$$

- (3) 有金标准的两组平行对照设计：
- 此时，主要衡量新诊断产品是否优于或者非劣于对照产品。譬如，通过做临床试验验证 3D 乳腺 X 射线的灵敏度和特异度是否优于 2D 乳腺射线。

		金标准		
新诊断产品		有病	无病	
	阳性	a_1	b_1	$a_1 + b_1$
	阴性	c_1	d_1	$c_1 + d_1$
		$a_1 + c_1$	$b_1 + d_1$	$a_1 + b_1 + c_1 + d_1$
对照产品		有病	无病	
	阳性	a_2	b_2	$a_2 + b_2$
	阴性	c_2	d_2	$c_2 + d_2$
		$a_2 + c_2$	$b_2 + d_2$	$a_2 + b_2 + c_2 + d_2$

新诊断产品的灵敏度= $\frac{a_1}{a_1+c_1}$

对照产品的灵敏度= $\frac{a_2}{a_2+c_2}$

若要比较新诊断产品的灵敏度是否优于对照产品，可以采用卡方检验或者近似正态 Z 检验。

新诊断产品的特异度 $=\frac{d_1}{b_1+d_1}$

对照产品的特异度 $=\frac{d_2}{b_2+d_2}$

若要比较新诊断的特异度是否优于对照产品，可以采用卡方检验或者近似正态 Z 检验。

(4) 无金标准时的两组配对设计：
对比新诊断产品和对照产品用 PPA 和 NPA 衡量。

		对照产品		
新诊断产品		阳性	阴性	
	阳性	<i>a</i>	<i>b</i>	<i>a + b</i>
	阴性	<i>c</i>	<i>d</i>	<i>c + d</i>
		<i>a + c</i>	<i>b + d</i>	<i>a + b + c + d</i>

PPA $=\frac{a}{a+c}$

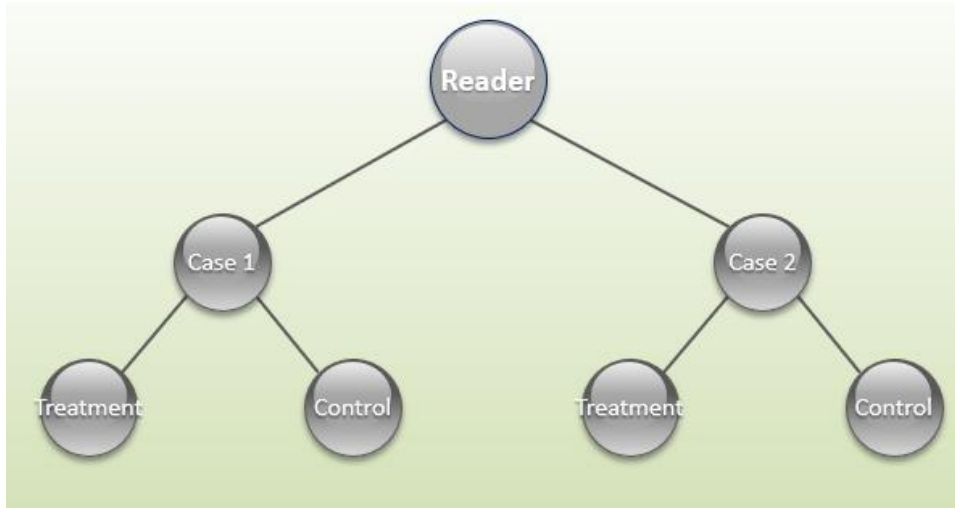
NPA $=\frac{d}{b+d}$

可以采用 McNemar 检验对新诊断和对照产品的阳性比例/阴性比例是否有差异做检验。

影像类诊断器械一般根据是否采用多阅片人，可以分为以下三种类型（以配对设计为例说明，关于平行对照设计可以思考下采用什么分析方法）：

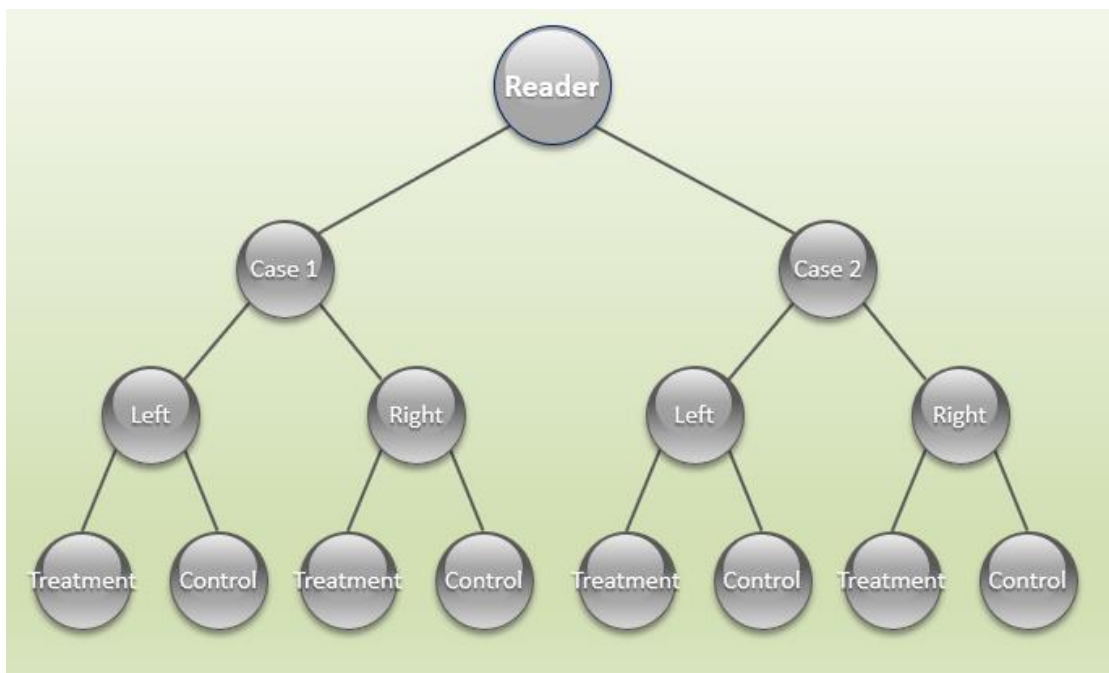
1. Matched-Pair design

不考虑阅片人之间的经验差异，用 1 个阅片人（实际中，往往采用两个阅片人+一个裁决者阅片；当两个阅片人判断结果不一致时，裁决者来判断结果是阳性还是阴性）采用新诊断产品和对照产品分别判断影像结果。如下图，阅片人对多个受试者分别采用新诊断产品（treatment）和对照产品（control）产生影像结果并进行判断。统计方法采用 McNemar 检验或 GEE 模型。



2. Clustered Matched-Pair design

不考虑阅片人之间的经验差异，用 1 个阅片人采用新诊断产品和对照产品分别判断影像结果，但是考虑到 cluster 结构，譬如，乳腺 X 射线判断乳腺结果时考虑到左右乳房的相关性，此时采用 clustered Matched pair 设计。统计分析方法采用 clustered McNemar 检验或 GEE 模型。



3. Multiple reader multiple case design 考虑阅片人之间的经验差异，用多个（例如，5 个阅片人）阅片人采用新诊断产品和对照产品分别判断影像结果。如下图，多个阅片人对多个受试者分别采用新诊断产品（treatment）和对照产品（control）产生影像结果并进行判断。统计方法采用 MRMC 混合效应模型或者 GEE 模型。

根据试验目的选好试验设计类型后，接下来重要的一步是计算样本量，那么对于有金标准的两组配对设计，样本量怎么计算呢？

诊断试验中有金标准的配对设计主要衡量新诊断产品的诊断准确率（灵敏度，特异度、ROC 曲线下面积）是否优于或者非劣于对照产品。譬如，通过做临床试验验证 3D 乳腺 X 射线的灵敏度和特异度是否优于 2D 乳腺射线。关于配对设计，需要考虑在产品之间诊断结果的相关性，然后计算样本量；如果不通过相关性来计算样本量，也可以通过四格表中的 a, b, c, d 来计算样本量。下面介绍下常见的几种样本量计算方法，并且对比这几种方法在相关性从 0 到 1 变化时的表现。配对设计中，我们对疾病受试者总数和健康受试者总数分别计算。本文以疾病受试者样本量为例，介绍常见的配对设计样本量计算方法。

已经过金标准确诊为疾病的受试者				
		对照产品		
新诊断产品		阳性	阴性	
	阳性	a_1	b_1	$a_1 + b_1$
	阴性	c_1	d_1	$c_1 + d_1$
		$a_1 + c_1$	$b_1 + d_1$	$a_1 + b_1 + c_1 + d_1$

已经过金标准确诊为疾病的受试者				
		对照产品 Y0		
新诊断产品 Y1		阳性	阴性	
	阳性	p_{11}	p_{10}	p_1
	阴性	p_{01}	p_{00}	$1 - p_1$
		p_0	$1 - p_0$	1

其中，

$$n = a_1 + b_1 + c_1 + d_1$$

$$p_{11} = \frac{a_1}{n}$$

$$p_{10} = \frac{b_1}{n}$$

$$p_{01} = \frac{c_1}{n}$$

$$p_{00} = \frac{d_1}{n}$$

$$\text{新产品的灵敏度} = p_1 = \frac{a_1 + b_1}{n}$$

$$\text{对照产品的灵敏度} = p_0 = \frac{a_1 + c_1}{n}$$

新诊断产品和对照产品之间的相关性大小：

$$cov(Y_1, Y_0) = \frac{p_1 + p_0 - 2p_1p_0 - p_{10} - p_{01}}{2} = p_{11} - p_1p_0$$

$$var(Y_1) = p_1(1 - p_1)$$

$$var(Y_0) = p_0(1 - p_0)$$

$$\rho = \frac{cov(Y_1, Y_0)}{\sqrt{var(Y_1)var(Y_0)}} = \frac{p_1 + p_0 - 2p_1p_0 - p_{10} - p_{01}}{2\sqrt{(p_1(1-p_1)p_0(1-p_0))}} = \frac{p_{11} - p_1p_0}{\sqrt{(p_1(1-p_1)p_0(1-p_0))}}$$

关于 $cov(Y_1, Y_0)$ 的具体推导参见 (Conner 1987, appendix)。

Miettinen 样本量公式 (Miettinen 1968):

$$\frac{[Z_{1-\alpha/2}\sqrt{p_{01} + p_{10}} + Z_{1-\beta}\sqrt{p_{01} + p_{10} - \frac{(p_1 - p_0)^2(3 + p_{01} + p_{10})}{4(p_{01} + p_{10})}}]^2}{(p_1 - p_0)^2}$$

Conner 样本量公式 (Conner 1987):

$$\frac{[Z_{1-\alpha/2}\sqrt{p_{01} + p_{10}} + Z_{1-\beta}\sqrt{p_{01} + p_{10} - (p_1 - p_0)^2}]^2}{(p_1 - p_0)^2}$$

独立两样本率比较的样本量公式 (药监局-医疗器械临床试验设计指导原则):

$$\frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2[p_1(1 - p_1) + p_0(1 - p_0)]}{(p_1 - p_0)^2}$$

GEE (logit function)样本量公式(Zhang 2014):

$$\frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2[p_1(1 - p_1) + p_0(1 - p_0) - 2\rho\sqrt{p_0(1 - p_0)p_1(1 - p_1)}]}{[p_0(1 - p_0)p_1(1 - p_1)](\log \frac{p_1}{(1 - p_1)} - \log \frac{p_0}{(1 - p_0)})^2}$$

GEE (identity function) 样本量公式 (Liu 1997):

$$\frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 [p_1(1-p_1) + p_0(1-p_0) - 2\rho\sqrt{p_0(1-p_0)p_1(1-p_1)}]}{(p_1 - p_0)^2}$$

GLMM (Generalized linear mixed model)样本量计算公式(Dang 2008):

$$N = \frac{VAR(\hat{\beta})(Z_{\alpha/2} - Z_{1-\Delta})^2}{h^2}$$

$$VAR(\hat{\beta}) = \left(\sum_{N_0} \frac{n_i}{1+(n_i-1)\rho_0^*\sigma_0^2} + \sum_{N_1} \frac{n_i}{1+(n_i-1)\rho_1^*\sigma_1^2} \right)^{-1} \left[\frac{1}{\pi\rho_0(1-p_0)} + \frac{1}{(1-\pi)p_1(1-p_1)} \right].$$

表 1 相关系数对样本量的配对率比较的样本量影响

一类错误/ 二类错误	p ₁	p ₀	p ₁₀	p ₀₁	ρ	Miettinen	Conner	GEE (Identity/logit)	独立两样本
0.05 / 0.2	0.9	0.8	0.1	0	0.667	56	77	71 / 75	197
			0.11	0.01	0.583	77	92	87 / 92	197
			0.12	0.02	0.5	96	108	103 / 108	197
			0.13	0.03	0.417	114	124	118 / 125	197
			0.14	0.04	0.333	131	139	134 / 141	197
			0.15	0.05	0.250	148	155	150 / 158	197
			0.16	0.06	0.1667	164	171	165 / 175	197
			0.17	0.07	0.083	181	186	181 / 191	197
			0.18	0.08	0	197	202	197 / 208	197
			0.19	0.09	-0.083	213	218	212 / 224	197
			0.2	0.1	-0.167	229	234	228 / 241	197

通过该表可以发现：

- Conner 方法算得的样本量大于 Miettinen 方法算得的样本量
- Conner 方法算得的样本量跟 GEE (logit)方法算得的样本量接近。
- 当相关性低于 0.5 时， GEE (identity)方法算得的样本量跟 Miettine 方法算得的样本量接近。

- GEE (logit)方法算得的样本量大于 GEE (identity)算得的样本量
- 当相关性大于 0.1 时，独立两样本方法算得的样本量最大。但是，当相关性小于 0.1 尤其是小于 0 时，独立两样本方法并不保守，算得的样本量小于其他几种方法。

到底选择哪个样本量公式，以下提供几点建议：

- 在试验设计阶段，通过模拟比较或者理论方法比较选择合适的分析方法，譬如到底是选择 GEE 模型、还是 GLMM 模型还是 McNemar 检验，可以在设计阶段通过模拟或者理论比较来选择。然后根据选定的方法选择相应的样本量估计方法。譬如，如果选择 GEE(identity)分析方法，那么就选择对应的样本量估计方法；如果选择 McNemar 分析方法，那么选择对应的比较保守的 Conner 方法。
- 根据以往数据来估计相关性的 大小，在相关性大于 0 的时候，选择保守的独立两样本方法估计样本量也是可以的。

参考文献：

1. Miettinen OS. The matched pairs design in the case of all-or-none responses. *Biometrics*. 1968;24 (2): 339–352.
2. Connor RJ. Sample size for testing differences in proportions for the paired-sample design. *Biometrics*. 1987; 43 (1): 207–211.
3. Zhang S, Cao J, Ahn C. A GEE approach to determine sample size for pre- and post-intervention experiments with dropout. *Computational Statistics and Data Analysis*. 2014; 69: 114–121.
4. Liu G, Liang KY. Sample size calculations for studies with correlated observations. *Biometrics*. 1997; 53: 937–947.
5. Dang TY, Mazumdar S, Houck PR. Sample Size and Power Calculations Based on Generalized Linear Mixed Models with Correlated Binary Outcomes. *Computer methods programs biomedicine*. 2008; 91(2): 122-127.
6. 医疗器械临床试验设计指导原则

