

样本量相关问题及 伦理考虑

康晓平

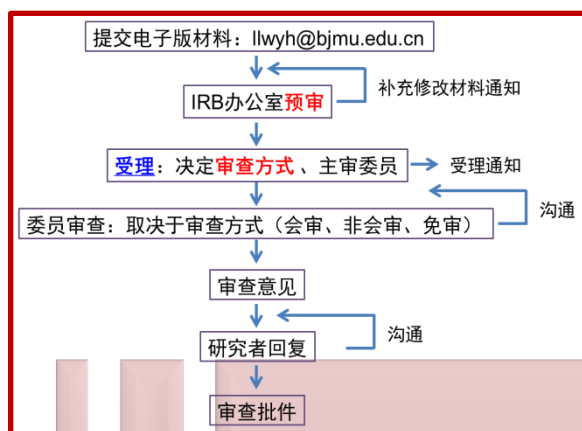
北京大学公共卫生学院

张海洪

北京大学受试者保护体系

2019.06.18

伦理审查的“生命周期”



初审

跟踪审查: 持续审查、修正案、不依从/方案违背、严重不良事件/非预期问题, 暂停或中止.....

结题

北京大学生物医学伦理委员会 (PU IRB)

初始审查申请表

七、研究摘要

研究设计
(可多选)

☐ 病例对照研究 ☐ 队列研究 ☐ 横断面研究
☐ 非随机对照研究 ☐ 随机对照研究 ☐ 应用盲法
☐ 其他:

研究背景和目的
(100字以内)

受试者选择
(80字以内)

对照设置、干预
措施、观察指标、
随访情况
(200字以内)

**样本量、
统计分析**

风险/受益分析
(50字以内)

拟研究时间

年 月 日至 年 月 日

北京大学生物医学伦理委员会 (PU IRB)

修正案审查申请表

三、修正情况

提出修正者：☐ 项目资助方 ☐ 研究中心 ☐ 项目负责人

修正类别：

☐ 研究设计 ☐ 研究步骤 ☐ 受试者例数 ☐ 纳入排除标准
☐ 干预措施 ☐ 知情同意书 ☐ 招募材料
☐ 其他（请说明）：

之前是否提请过修正案：☐ 是 次数__
☐ 否

内容

- 与样本量有关的主要参数
及常用的样本量估计方法
- 与样本量有关的伦理考虑

-- 样本量的主要参数

1. 检验水准 $\alpha=0.05$

2. 检验效能 (power) = $1-\beta$ 为 0.9 或 0.8

3. 标准差 S (σ 的估计值)

4. 两个总体参数间的差值 $\delta = \mu_1 - \mu_2$, 或
 $\delta = \pi_1 - \pi_2$

S 、 δ 这两个参数可来自预实验、相关文献、
或专业领域认可的经验值

常用的样本量计算公式（1）

研究方法	样本量计算公式
病例对照 （因素探讨）	$n = \frac{(Z_{\alpha/2} \sqrt{2pq} + Z_{\beta} \sqrt{p_0 q_0 + p_1 q_1})^2}{(p_1 - p_0)^2}$
队列研究 （因素探讨）	同上
两组平行对照 的临床试验 （差异性研究）	$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\delta^2} [p_1(1-p_1) + p_2(1-p_2)]$
两生存率比较 （随访研究）	$n = \{Z_{\alpha}[p^2(Q_e^{-1} + Q_c^{-1})]^{1/2} + Z_{\beta}(p_e^2 Q_e^{-1} + p_c^2 Q_c^{-1})^{1/2}\}^2 / (p_e - p_c)^2$

常用的样本量计算公式（2）

研究方法		样本量计算公式
非劣性试验 (证实性研究)	定量指标	$n = 2\left[\frac{(Z_{\alpha} + Z_{\beta})s}{\delta}\right]^2$
	率指标	$n = 2\pi(1 - \pi)\left[\frac{Z_{\alpha} + Z_{\beta}}{\delta}\right]^2$
等效性试验 (证实性研究)	定量指标	$n = 2\left[\frac{(Z_{\alpha} + Z_{\beta/2})s}{\delta}\right]^2$
	率指标	$n = 2\pi(1 - \pi)\left[\frac{Z_{\alpha} + Z_{\beta/2}}{\delta}\right]^2$

1、病例对照样本量计算

成组资料比较且人数相等时，算法应用以下公式：

$$\bullet n = \frac{(Z_{\alpha/2}\sqrt{2pq} + Z_{\beta}\sqrt{p_0q_0 + p_1q_1})^2}{(p_1 - p_0)^2} \quad (1)$$

公式（1）为双侧检验用；若为单侧检验，则式中的 $Z_{\alpha/2}$ 换成 Z_{α} 。式中 n 为所需的病例组和对照组人数， $Z_{\alpha/2}$ 或 Z_{α} 和 Z_{β} 分别与 α 和 β 取值对应的 Z 界值。 p_0 与 p_1 分别为对照组与病例组估计的某因素暴露史的比例。

$q_0 = 1 - p_0, q_1 = 1 - p_1, p = (p_0 + p_1)/2, q = 1 - p$, 其 p_1 可用 OR 值推出

$$\bullet p_1 = (OR \times p_0) / (1 - p_0 + OR \times p_0) \quad (2)$$

例1:

- 在一项肺癌与吸烟的研究中，已知对照人群吸烟率为30%，预计吸烟的 $OR=3.0$ ，取 $\alpha=0.05$ ， $\beta=0.10$ 。试估计病例组与对照组的例数？

- 已知 $p_0=0.30$ ， $q_0=1-p_0=0.70$ ， $OR=3.0$ ， $Z_{0.05/2}=1.96$ ， $Z_{0.1}=1.28$

$$p_1=(0.30)(3.0)/[1-0.30+(3.0 \times 0.30)]=0.5625$$

$$q_1=1-0.5625=0.4375$$

$$p=(0.5625+0.30)/2=0.4313$$

$$q=1-0.4313=0.5687$$

$$n=\frac{(1.96\sqrt{2 \times 0.4313 \times 0.5687} + 1.28\sqrt{0.30 \times 0.70 + 0.5625 \times 0.4375})^2}{(0.5625 - 0.30)^2} = 72.6 \approx 73 \text{ (人)}$$

本研究的病例组和对照组各需要73人。

2、队列研究的样本量计算

- 队列研究仍然可用公式（1）计算，但这里 p_1 为暴露组的发病率， p_0 为非暴露组的发病率，其余均同病例对照研究。

$$n = \frac{(Z_{\alpha/2} \sqrt{2pq} + Z_{\beta} \sqrt{p_0 q_0 + p_1 q_1})^2}{(p_1 - p_0)^2} \quad (1)$$

- 例2:

某研究者采用队列研究探讨孕妇服用某药与婴儿先天性心脏病的关系。文献报道，未服用此药先天性心脏病的发病率为1%。估计服用该药的RR为2.0，设 $\alpha=0.05$ ， $\beta=0.10$ ，试计算该研究的样本量。

已知 $p_0=0.01$ ， $q_0=0.99$ ， $p_1=2.0 \times 0.01=0.02$ ， $q=0.98$

$p=(0.01+0.02)/2=0.015$ ， $q=0.985$

$$n = \frac{(1.96 \sqrt{2 \times 0.015 \times 0.985} + 1.28 \sqrt{0.01 \times 0.99 + 0.02 \times 0.98})^2}{(0.02 - 0.01)^2} = 3100 \text{ (人)}$$

本研究的暴露组和非暴露组各需要3100人。

3、两组平行对照的临床试验 (差异性检验)

$$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\delta^2} [p_1(1 - p_1) + p_2(1 - p_2)] \quad (2)$$

- p_1 为第一组的率, p_2 为第二组的率, δ 为具有临床意义的两组差值; 比较的两组病例数相等, 双侧检验。

4、两生存率比较样本含量的估计

- 估计公式为:

$$n = \{Z_{\alpha}[p^2(Q_e^{-1} + Q_c^{-1})]^{1/2} + Z_{\beta}(p_e^2 Q_e^{-1} + p_c^2 Q_c^{-1})^{1/2}\}^2 / (p_e - p_c)^2 \quad (3)$$

式中 p_e 、 p_c 为试验组、对照组事件发生率, Q_e 、 Q_c 分别为试验组、对照组样本含量构成比($Q_e + Q_c = 1$); 各组样本含量: $n_e = nQ_e$, $n_c = nQ_c$; α 有单、双侧之分, Z_{α} 、 Z_{β} 查Z界值表, $p = p_e Q_e + p_c Q_c$ 。

例3：

某医师，欲比较两种早期乳腺癌疗法的治疗效果，已知旧疗法5年生存率为52.5%，新疗法5年生存率为72.5%，当 $\alpha=0.05$ ， $\beta=0.10$ 时要观察多大的样本量？

本例单侧检验， $p_e=0.725$ ， $p_c=0.525$ ， $Q_e=Q_c=0.5$ ，

$p=(0.5)(0.725)+(0.5)(0.525)=0.625$ ，查Z界值表 $Z_{0.05}=1.6449$ ， $Z_{0.10}=1.2816$ ，则

$$n = \frac{\{(1.6449)[(0.625)^2(0.5^{-1} + 0.5^{-1})]^{\frac{1}{2}} + (1.2816)[(0.625)^2 0.5^{-1} + (0.525)^2 0.5^{-1}]^{1/2}\}^2}{(0.725 - 0.525)^2}$$

≈ 339

$$n_e = n_c = 0.5 \times 339 \approx 170$$

这项设计新、旧疗法各需要观早期乳腺癌患者170例。

5、非劣性试验

- 5.1 定量指标

非劣性试验应当采用单侧的检验水准 α ，假定允许的二类误差概率不超过 β ，则在2组总体均数相同，即 $T = S$ 的条件下，非劣性试验每组需要的样本含量为：

- $$n = 2 \left[\frac{(Z_{\alpha} + Z_{\beta})s}{\delta} \right]^2 \quad (4)$$

式中 Z 为标准正态离差界值， Z_{α} 和 Z_{β} 均为单侧界值。 s 为2组的合并标准差。 n 为每组的样本含量。

例4：

为了对AII与ACEI疗效进行比较,确认AII是否不比ACEI的疗效差,假定AII与ACEI的降压作用相同,若事先确定 $\delta = 0.40\text{kPa}$,已知2组合并标准差 $s = 1.06\text{kPa}$,按单侧 $\alpha = 0.05$,试计算得出非劣效结论有90%把握度(即 $\beta = 0.10$)时的每组样本含量。

显然,本例试验为非劣性设计,代入有关量到公式(4),可求得每组的含量为:

$$n = 2 \left[\frac{(1.645 + 1.282) \times 1.06}{0.40} \right]^2 = 120.3 \approx 120$$

即每组用120例患者进行试验,即可有90%的把握度,在 $\alpha = 0.05$ 的显著性水平、非劣性界值为 0.40kPa 的情况下得出非劣效结论。

5、非劣性试验

- 5.2 率指标

非劣性试验应当采用单侧的检验水准 α ，假定允许的二类误差概率不超过 β ，则在2组总体率相同，即 $T = S = \pi$ 的条件下，非劣性试验每组需要的样本含量为：

$$n = 2\pi(1 - \pi) \left[\frac{Z_{\alpha} + Z_{\beta}}{\delta} \right]^2 \quad (5)$$

例5：

治疗手足癣常规推荐伊曲康唑100mg 用药4周方案，考虑到该药有极高的组织亲和力，停药后可在角质层持续停留4周，这种药动学特性提示服药1周的短程疗法可能对皮肤真菌病同样有效。为此设计了400mg ×1周和100mg ×4周的比较试验，考察新方案的疗效是否不比常规方案差。以临床治愈率作为评价终点,假设两方案的疗效相同，根据以前的疗效及有关要求,取： $T = S = \pi = 0.80$ ， $\delta = 0.15$ ， $\alpha = 0.05$ ， $\beta = 0.20$ 。则单侧 $Z_{0.05} = 1.645$ ，单侧 $Z_{0.20} = 0.845$ ，每组需要的样本含量按公式(5) 计算如下：

$$n = 2 \times 0.80(1 - 0.80) \left[\frac{1.645 + 0.845}{0.15} \right]^2 = 88.2 \approx 88$$

研究方案已确定的样本量在研究实 施过程中是否可以修改？

- 原则：

- 科学依据

- 伦理依据

北京大学生物医学伦理委员会（PU IRB）

修正案审查申请表

三、修正情况

提出修正者：☐ 项目资助方 ☒ 研究中心 ☐ 项目负责人

修正类别：

☒ 研究设计 ☐ 研究步骤 ☒ 受试者例数 ☒ 纳入排除标准

☐ 干预措施 ☒ 知情同意书 ☐ 招募材料 ☐

其他（请说明）：

之前是否提请过修正案：☐ 是 次数__

☒ 否

为了避免对受试者造成紧急伤害，是否在提交伦理委员会审查批准前对方案进行了修改并实施：

☐ 是 ☒ 否 ☐ 不适用

修正的具体内容与原因：（也可另附文件说明）

见附件

北京大学生物医学伦理委员会（PU IRB）

研究进展报告

1. 受试者数目		
计划入组例数	轻症xxx病例40例， 重症xxx病例20例	
是否准备更改受试者人数	<input checked="" type="checkbox"/> 是（参见不依从事件报告） <input type="checkbox"/> 否	
1. 目前受试者入组信息汇总		
	过去一年（或上次持续审查至今）	累计数
1) 入组的受试者	228	228
1) 入选的潜在弱势群体数量	228	228
1) 已签署知情同意书、但自动退出的受试者数量	0	0
1) 已签署知情同意书、但失去联系的受试者数量	0	0
1) 已签署知情同意书、但被剔除的受试者数量	0	0
1) 已签署知情同意书并完成研究的受试者数量 （所有干预及随访都已完成）	228	228

研究人数、范围或规模是否已发生变化，或预期在未来一年内会发生变化，而这些变化将需要其他资源？

否

☒ 是。如果是，请说明：因为xxx病疫情较严重，为了更好的研究xxx病的发病机制和重症转归机制，我们扩大了采样的规模，详情见不依从事件报告。

北京大学生物医学伦理委员会（PU IRB）

不依从事件报告表

二、不依从事件情况与原因（请说明发生日期 / 发现日期 / 发现途径等）

原计划将40例轻症xxx病例和20例重症xxx病例纳入研究，但随着样本采集的深入，xxx病疫情于2014年夏季有所爆发，为了更好的研究xxx病的发病机制和重症转归机制，我们适当扩大了采样的规模，共收集143例轻症病例和85例重症病例。

- 是否想过为什么要估计样本量？

思考：

-----科学方面

-----伦理方面

Under-sampled Studies are more Harmful to Study Participants than Over-sampled Studies

Thomas E. Wasser, PhD, MEd, CIM
HealthCore, Inc., Wilmington, DE, USA

BACKGROUND

The concept of equipoise is known by ethics professionals. Under-sampled studies have the potential for more harm to participants than studies that are over-sampled.

Figure 1. Correctly sampled study. Equipoise assumed to be present.



Figure 2. Equipoise in an under-sampled study. IRB should reject this study.



This research demonstrates why under-sampling carries more risk for harm and presents methods that IRBs can use to identify and correct studies where under-sampling can occur.

METHODS

Using pretest-posttest control group design, a power analysis determines that $n=100$ participants are needed in each group ($n=200$ total). In scenario 1, a study enrolls 80% of the total and in scenario 2, a study enrolls 120% of the total. In the under-sampled study, all 80 patients have been exposed to risk with no study benefit. In the over-sampled study, the study had benefit and only the 20 additional participants were exposed to risk.

CONCLUSION

In minimal or high-risk studies where low enrollment might be anticipated, there are several procedures that hospital and academic IRBs can use to detect and remedy under-sampled studies. The following recommendations can be added to investigator questionnaires and included in either proxy or direct IRB presentations:

Table 1. Decision matrix for IRB when sampling issues are discussed.

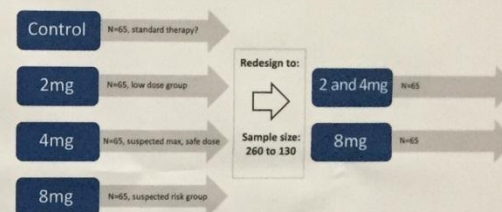
IRB and Protocol Definitions (Assumes Minimal Risk or Greater)		Protocol contains		
		Under-Sample	Appropriate	Over-sample
IRB Determines the following Exists	Under-sample	Reject	Table Ask for revision or clarification	Reject
	Appropriate	Table Ask for revision or clarification	Approve	Approve but Monitor
	Over-sample	Reject	Approve but Monitor	Approve but Monitor

1. Ensure that all proposals submitted to the IRB have sample size/power sections indicating the number of participants needed for the study.
2. Confirm that the data analysis section of the proposal matches the sample size calculations and are appropriate for the statistics being used.
3. Require that the presenter (or IRB assigned/proxy presenter) be familiar with the sample size requirements for the study and can discuss them accurately.
4. Add questions regarding participant recruitment procedures.
5. Ask the presenter (or query an existing EMR system) to determine if the presenter has access to the patient population needed to complete the enrollment.
6. If the IRB is concerned that the study enrollment is problematic, require the presenter to provide more than annual reports on study recruitment. The IRB could require monthly reporting or reporting after each patient is enrolled.

7. Appoint IRB members or external members to monitor study or site enrollment and make recommendations back to the IRB.
8. If the study is multi-centered, base approval of the study contingent upon the sponsor providing data regarding these recommendations.

While this example included only one type of analytical design—t-tests for between group or within group comparisons. Every statistic has its own unique power formula, so risk to participants would need to be computed for each type of analytical procedure. IRBs should not hesitate to provide research design suggestions to improve sample size selection.

Figure 3. Method for IRB to suggest a research design modification to protect against over sampling



DISCUSSION

Research protocols should be closely examined for the effects of under and over sampling. Procedures or a checklist should be put in place by the IRB to continually monitor, on more than an annual basis, those studies that are not enrolling subjects as specified in the protocol. The steps for ensuring proper sampling procedures can be assembled into a Decision Tree to help facilitate this process.

HealthCore

Poster presented at PRIM&R Conference, San Diego, California, November 14–17, 2018.

Poster: **Under-sampled** studies are **more harmful** to study participants than **Over-sampled** studies

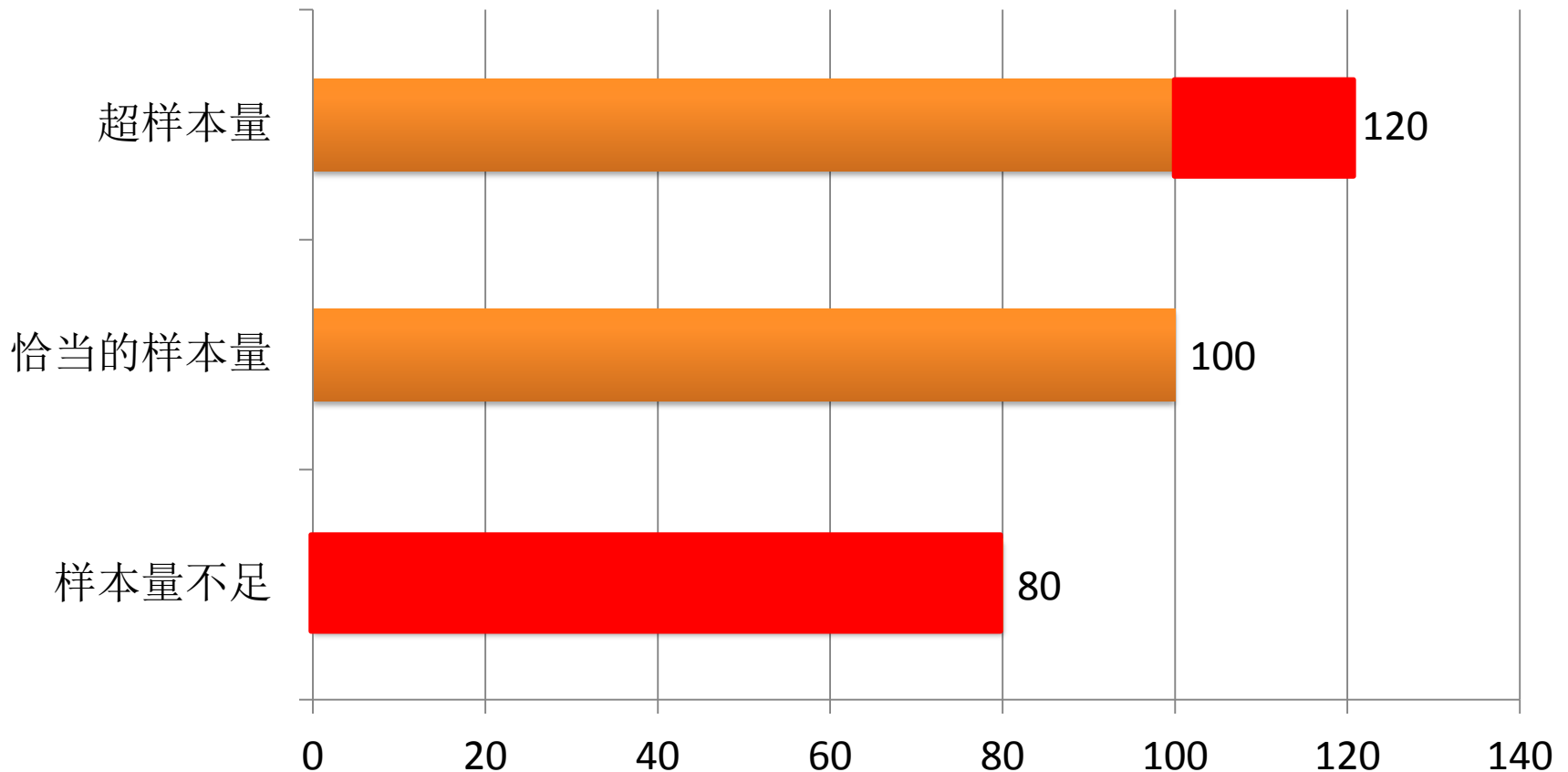
背景:

- 临床均势（ **equipoise**）——伦理角度
- 样本量不足对受试者造成的潜在风险更大

目的:

- 解释为什么样本量不足造成的风险更大
- **IRB**如何识别和管理那些可能存在样本量不足这一问题的研究

Poster: **Under-sampled** studies are **more harmful** to study participants than **Over-sampled** studies



样本量 vs. 伦理考虑

- 研究样本量的决定因素？
 - 统计学依据
 - 成本/预算
 - 受试者的参加意愿，等



样本量 vs. 伦理考虑

- 研究方案

- 样本量及其计算依据
- 数据分析
- 受试者招募

- 伦理审查

- 资源/资质
- 样本量是否合理
- 招募计划/材料是否合适

样本量 VS. 伦理考虑

- 研究实施

- 受试者招募

- 进展
 - 人群
 - 样本量

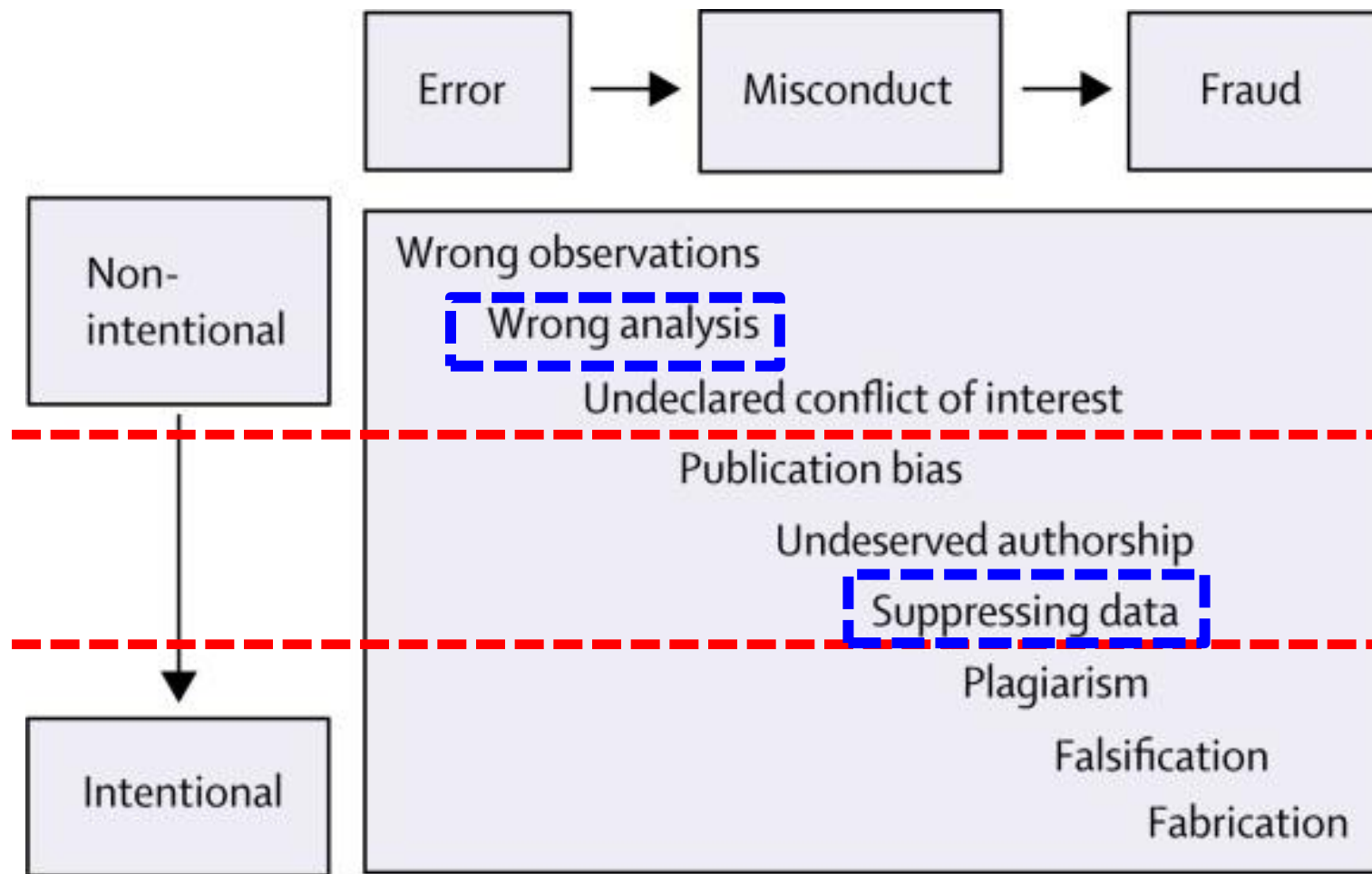
- 伦理审查

- 年度持续/跟踪审查
 - 修正案审查
 - 方案违背

举例&讨论

- 研究者开展一项功能食品试食试验，在伦理申请表中声明“无项目资助”
- 一项临床研究计划在3年内纳入100例受试者，第二年持续审查时，共计纳入受试者15例
- 一项流行病学调查研究（不大于最小风险）原计划纳入800名受试者，实际纳入1800例

拓展：数据分析与样本量（伦理角度）



Magne Nylenna, Sigmund Simonsen, Scientific misconduct: a new approach to prevention, In The Lancet, Volume 367, Issue 9526, 2006, Pages 1882-1884.

谢谢！

问题与讨论

