—— 计算题　—— 概述题　—— 解答

Clinical Trial

2024

算人数：$N = \frac{\theta_A(1-\theta_A)+\theta_B(1-\theta_B)}{(\theta_A-\theta_B)^2}\left[\Phi^{-1}(1-\frac{\alpha}{2})+\Phi(1-\beta)\right]^2$

最后总共需要 2N 人（向上取整）

实践中人不够？患者可能退出 (withdraw)，估计退出率，

相应扩大规模.

Meta 分析的四个主要目标：① 一致、客观地展示不同试验的数据

② 测试总体零假设

③ 估计平均治疗效果

④ 调查试验之间是否存在统计异质性. heterogeneity

Meta 分析中 Fixed effect：$\hat{\mu}_{FE} = \frac{\sum W_c \hat{Y}_c}{\sum W_c}$，where $W_c = \frac{1}{Var[\hat{Y}_c]}$

$\Rightarrow Var[\hat{\mu}_{FE}] = \frac{1}{\sum W_c}$

$\Rightarrow$ 95% CI：$\hat{\mu}_{FE} \pm 1.96\sqrt{Var[\hat{\mu}_{FE}]}$ 含 0 不显著.

Publication bias 发表偏差：指大型或显著治疗效应会发表在医学期刊上，那么较小或非显著影响更容易被看到的现象.

通俗来讲. 100个团队只有5个成功了发表，而另外95个都失败了 —— "管中窥豹"

即并非真相. 而是被筛选后的真相

判断依据：漏斗图 (Funnel plot) $\begin{cases} \text{对称} \to \text{无 bias} & \text{symmetric} \\ \text{不对称} \to \text{有 bias} & \text{asymmetric} \end{cases}$

$\Rightarrow$ 缺点：图形诊断 (graphical diagnostic)，解释并不清晰 (interpretation)

证明 Random Effect：if $\hat{Y}_c \sim N(\mu, \frac{1}{W_c}+\tau^2)$.　$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

$$\Rightarrow l(\mu|\tau^2) = \log L(\mu|\tau^2) = \log\left[\prod_{c=1}^{C}\frac{1}{\sqrt{2\pi(\frac{1}{W_c}+\tau^2)}}\exp\left(-\frac{(\hat{Y}_c-\mu)^2}{2(\frac{1}{W_c}+\tau^2)}\right)\right]$$

$$= \sum_{c=1}^{C}\left(-\frac{1}{2}\log\left(2\pi\left(\frac{1}{W_c}+\tau^2\right)\right) - \frac{(\hat{Y}_c-\mu)^2}{2(\frac{1}{W_c}+\tau^2)}\right)$$

$$\Rightarrow \frac{d}{d\mu}l(\mu|\tau^2) = \sum_{c=1}^{C}\frac{2(\hat{Y}_c-\mu)}{2(\frac{1}{W_c}+\tau^2)} = \sum_{c=1}^{C}\frac{\hat{Y}_c-\mu}{\frac{1}{W_c}-\tau^2} = 0$$

$$\Rightarrow \hat{\mu}_{RE} = \frac{\sum_{c=1}^{C}\frac{\hat{Y}_c}{1/W_c-\tau^2}}{\sum_{c=1}^{C}\frac{1}{1/W_c-\tau^2}}.$$

2023

Randomisation

Double-blind

Placebo-controlled

4 stages to design clinical trial

算人数 (连续).

$N = \frac{2\sigma^2}{(\mu_A-\mu_B)^2}\left[\Phi^{-1}(1-\frac{\alpha}{2})+\Phi(1-\beta)\right]^2$

Meta 分析：应用统计方法于问题时结合结果 from 不同临床试验时同一个治疗方法

识别 Publication Bias

95% CI of odd ratio 含 1 不显著

Odd ratio 是两个 odd 之比.

均为失败/成功 or desease / no desease

$\frac{\text{risk}}{\text{no risk}}$

2022

算人数 (离散)

Phrase IV

Control group?

Randomisation

OR

2021

Epidemiology

2024

$$\text{Sensitivity} = \frac{\text{Number of diseased people who screen positive}}{\text{Total number of diseased people}}$$

$$\text{Specificity} = \frac{\text{.. .. healthy .. .. .. negative}}{\text{Total number of healthy people}}$$

Direct Standardisation: 公平比较2个不同人口统计学特征的独立人群疾病率

$$\Rightarrow \text{Age standardised mortality rate (ASMR)} = \frac{\sum \frac{y_i}{n_i} \times N_i}{\sum N_i} \longrightarrow \text{Var}[\text{ASMR}] = \frac{\sum \frac{y_i}{n_i^2} \times N_i^2}{(\sum N_i)^2}$$

Assumption: $y_i \sim \text{Poisson}(n_i, \theta_i)$

$\Rightarrow \hat{\theta}_{MLE} = \frac{y_i}{n_i}$

Indirect Standardisation: 在参考群体中特定年龄死亡率下,计算研究群体死亡数.
然后在两个群体比较发病率

$$\Rightarrow \text{Standardised mortality rate (SMR)} = \frac{\text{Observed death}}{\text{Expected death}} = \frac{Y}{E} = \frac{\sum y_i}{\sum n_i \cdot r_i}$$

均表示study population死亡率
比 reference population
多(少) $|(A)SMR-1|$ 比例.

$\text{Var}[\text{SMR}] = \frac{Y}{E^2}$

95% CI: $\text{SMR} \pm 1.96\sqrt{\frac{Y}{E^2}}$ (含1不显著)

Assumption.
$Y \sim \text{Poisson}(ER)$.
$\hat{R}_{MLE} = \frac{Y}{E}$

Spatial ecological study: 基于小地理区域内的个体群体.
收集每个地理区域内居住人群的疾病总水平.
平均暴露水平和其他重要协变量数据
⇒ 通过比较不同空间单元内居住的群体,评估暴露对疾病在人群水平上的影响

还有很多. 但这些不要也罢.

2023

$$OR = \frac{\text{Disease / no Disease (at risk)}}{\text{Disease / no Disease (no risk)}} \Rightarrow \text{Var}[\log(OR)] = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \Rightarrow 95\%\text{ CI}: \log(OR) \pm 1.96\sqrt{\text{Var}} \Rightarrow (\exp(\cdot), \exp(\cdot)).$$

Mantel-Haenszel 方法. $OR_{MH} = \frac{\sum w_i OR_i}{\sum w_i}$, where $w_i = \frac{1}{\text{Var}[\log OR]} \approx \frac{b_i c_i}{n_i}$ (副对角线).

Sensitivity, Specificity.

2022

risk的disease率

no risk disease率

Attributable risk: $AR = P_1 - P_2 = \frac{A}{A+B} - \frac{C}{C+D}$.

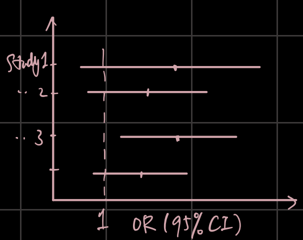
总患病数

Population attributable risk: $PAR = \frac{(A+C) - NP_2}{A+C}$ ← 总人数的预期值需要

Relative risk: $RR = \frac{P_1}{P_2} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}} \Rightarrow \text{Var}[\log(RR)] = \frac{1}{A} - \frac{1}{A+B} + \frac{1}{C} - \frac{1}{C+D}$.

2021

OR. OR_{MH}



Survival Analysis

2024

Kaplan-Meier 估计：$\hat{S}_{KM}(t) = \prod_{i=1}^{n} \frac{s_i}{r_i}$

$s_i = r_i - d_i$ 减去当天死的（不包括 censor）

r_i：t_i前还活着的（当天死的不算）

对于 $1^+, 2, 2, 5, 6, 6^+, 7, 9^+, 10^+, 12$. Here are 5 distict time: 2, 5, 6, 7, 12

Event time $t_{(i)}$	Number at risk at time $t_{(i)}$ r_i（从这个时间点往后查人数即可）	Number of surviving at time $t_{(i)}$ s_i	Estimate survival function at time $t_{(i)}$ $\hat{S}(t_{(i)})$
※ 0	10 一共10个人	10	1
2	9 1^+死了	$9-2=7$ $t=2$时死了2个	$7/9$
5	7 $1^+, 2, 2$死了	$7-1=6$ $t=5$时死了1个	$7/9 \times 6/7 = 2/3$
6	6	$6-1=5$ 6^+不算	$2/3 \times 5/6 = 5/9$
7	4	3	$5/9 \times 3/4 = 5/12$
12	1	0	0

那对于 $t=3$ 和 $t=4$ 都是还没到 $t=5$ 的时间. 所以都按之前的 $t=2$ 算，即 $7/9$.

⇒ 缺点：Kaplan Meier estimator is a step function. 这使得 $t=3$ 和 $t=4$ 一样. 用平滑器 (smoother)，使其不再是一个阶梯函数.

Assumption: $s_i \sim \text{Binomial}(r_i, \theta_i) \Rightarrow \hat{\theta}_{i\,MLE} = \frac{s_i}{r_i} \Rightarrow \Rightarrow \Rightarrow \text{Var}[\hat{S}_{KM}(t)] = (\hat{S}_{KM}(t))^2 \sum \frac{r_i - s_i}{r_i s_i} \Rightarrow$ 95% CI: $\hat{S}_{KM}(t) \pm 1.96\, \hat{S}_{KM}(t) \sqrt{\sum_{i=1}^{n} \frac{r_i - s_i}{r_i s_i}}$

Cox proportional hazards model (PHM):

$$h(t, z_i) = h_0(t) \exp(z_i^T \beta)$$

$$\exp\left(-\int_0^t h(u, z_i)\,du\right) = \exp\left(-\int_0^t h_0(u)\,du \cdot \exp(z_i^T\beta)\right) = \exp\left(-\int_0^t h_0(u)\,du\right)^{\exp(z_i^T\beta)}$$

$$\Rightarrow S(t, z_i) = S_0(t)^{\exp(z_i^T\beta)}$$

is the hazard function at time t for individual i with covariate vector $z_i = (z_{i1}, \cdots, z_{ip})$.
where $h_0(t)$ is baseline hazard function, β is a set of unknown parameters

⇒ Hazard ratio: $\exp(\hat{\beta}) \Rightarrow$ 95CI: $\hat{\beta} \pm 1.96 \cdot SE[\hat{\beta}] \Rightarrow \exp(\hat{\beta} \pm 1.96 SE[\hat{\beta}])$ 含1不显著

2023 基本一样.

① $h(t, z_1) = C \times h(t, z_2)$

$\Rightarrow H(t, z_1) = C \times H(t, z_2)$

$\Rightarrow -\log[S(t, z_1)] = C \times (-\log[S(t, z_2)])$

$\Rightarrow \log\{-\log[S(t, z_1)]\} = \log(C) + \log\{-\log[S(t, z_2)]\}$.

$\Rightarrow X_1(t) = \log\{-\log[S(t, z_1)]\} = \log[H(t, z_1)]$

$X_2(t) = \log\{-\log[S(t, z_2)]\} = \log[H(t, z_2)]$

parallel

② $\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \sim \chi_1^2(1-\alpha)$

$O_A = \sum d_{Ai} \quad E_A = \sum E_{Ai} = \sum r_{Ai} \frac{d_i}{r_i}$

A: 2, 3+, 5+, 7, 7, 11

B: 4, 5, 6, 8+, 9+, 11+

$t_{(i)}$	d_A	d_B	$d_A + d_B$	r_A	r_B	$r_A + r_B$	E_{Ai}	E_{Bi}
2	1	0	1	6	6	12	$6 \times 1/12$	$6 \times 1/12$
4	0	1	1	4	6	10	$4 \times 1/10$	$6 \times 1/10$
5	0	1	1	4	5	9	$4 \times 1/9$	$5 \times 1/9$
6	0	1	1	3	4	7	$3 \times 1/7$	$4 \times 1/7$
7	2	0	2	3	3	6	$3 \times 2/6$	$3 \times 2/6$
11	1	0	1	1	1	2	$1 \times 1/2$	$1 \times 1/2$
	$O_A = 4$	$O_B = 3$					$E_A =$	$E_B =$

基础公式：Survival function: $S(t) = 1 - F(t) = \exp\left(-\int_0^t h(u)\,du\right)$

hazard function: $h(t) = \lim_{\Delta t \to 0} \frac{\Pr(T \le t + \Delta t \mid T \ge t)}{\Delta t}$ (马上就死).

$= \frac{f(t)}{S(t)} \Rightarrow H(t) = \int_0^t h(u)\,du$

$f(t) = F'(t) = (1 - S(t))' = -S'(t)$

$h(t) = \frac{-S'(t)}{S(t)} = -(\log S(t))'$

$\Rightarrow \log S(t) = -\int_0^t h(u)\,du \quad \Rightarrow S(t) = \exp(-H(t)) \quad \Rightarrow H(t) = -\log(S(t))$

2022.

Joint likelihood: $L(t_1, \cdots, t_n, \delta_1, \cdots, \delta_n) = \prod_{i=1}^{n} \left\{ [\underbrace{f(t_i | z_i)}_{\text{未删}}]^{\delta_i} [\underbrace{S(t_i | z_i)}_{\text{删}}]^{1-\delta_i} \right\}$.

↓

log-likelihood: $\log L$

2021

Nelson and Aalen: $\hat{H}_{NA}(t) = \sum \frac{d_i}{r_i}$, $\chi^2 \begin{cases} H_0\text{: 生存分布(总数)一致} \\ H_1\text{: 生存分布(总数)不一致} \end{cases}$