256 [55]

P1

66

1. (i) In study one, if a person's unhealthy diet increase by 1 unit, the odds of developing CHD increses by 21.1%. ~~[illegible]~~ 95% confidence interval of the odds of developing CHD raging from decreasing 17.5% to increasing 77.8%. Since the 95% CI ~~[illegible]~~ includes 1, this means the effect is not statically significant. // 2/2

(ii)

$$OR_2 = \frac{\text{odds of failure on treatment}}{\text{odds of failure on control}} = \frac{\frac{45}{148}}{\frac{35}{221}} = \frac{221}{154} \approx 1.435$$

$$V(\log(OR_2)) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} = \frac{1}{45} + \frac{1}{35} + \frac{1}{158} + \frac{1}{224} \approx 0.0604$$ // ✓ 2/2

(iii)

$$W_i = \frac{b_i \times c_i}{n_i}$$, where ~~b_i is number of people with risk factor present and unhealthy, c_i is number of healthy people with risk factor absent. n_i is number of~~

where b_i is number of healthy people with risk factor present in group i,

c_i is number of diseased people with risk factor absent in group i,

n_i is number of all people being investigated in group i

2/2

(iv)

$$W_2 = \frac{198 \times 35}{499} = \frac{6930}{495} \approx 13.888$$ // ✓ 1/1

256155

P 2

Study	OR	V(log OR)	W
1	1.241	0.0384	23.7
2	1.435	0.0604	13.888
3	1.276	0.0255	33.193

(v) $OR_{MH} = \frac{\sum_{i=1}^{3} w_i \times OR_i}{\sum_{i=1}^{3} w_i}$

$OR_2 = 1.435$

$V(\log OR_2) = 0.0604$

$W_2 = 13.888$

$\Rightarrow OR_{MH} = \frac{23.7 \times 1.241 + 13.888 \times 1.435 + 33.193 \times 1.276}{23.7 + 13.888 + 33.193}$

$= 1.285$ ✓

2/2

(vi) $V[\log(OR_{MH})] = \frac{\sum_{i=1}^{3} w_i^2 V_i}{(\sum_{i=1}^{3} w_i)^2}$, where $w_i = \frac{b_i \times c_i}{n_i}$ (same as part iii)

$V_i = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}$

a_i: number of diseased people with risk factor present in group i

b_i: number of healthy people with risk factor present in group i

c_i: number of diseased people with risk factor absent in group i ✓

d_i: number of healthy people with risk factor absent in group i.

h_i: number of people being investigated in group i.

3/3

(vii)

95% CI for OR_{MH}: $\exp[\log(OR_{MH}) \mp 1.96\sqrt{V(\log(OR_{MH}))}]$

$V[\log(OR_{MH})] = \frac{23.7^2 \times 0.0384 + 13.888^2 \times 0.0604 + 33.193^2 \times 0.0255}{[23.7 + 13.888 + 33.193]^2} = 0.0122$ ✓

$\Rightarrow$ 95% CI for OR_{MH}: $(\exp(\log(1.285) \mp 1.96 \times \sqrt{0.0122}))$ ✓

1.115

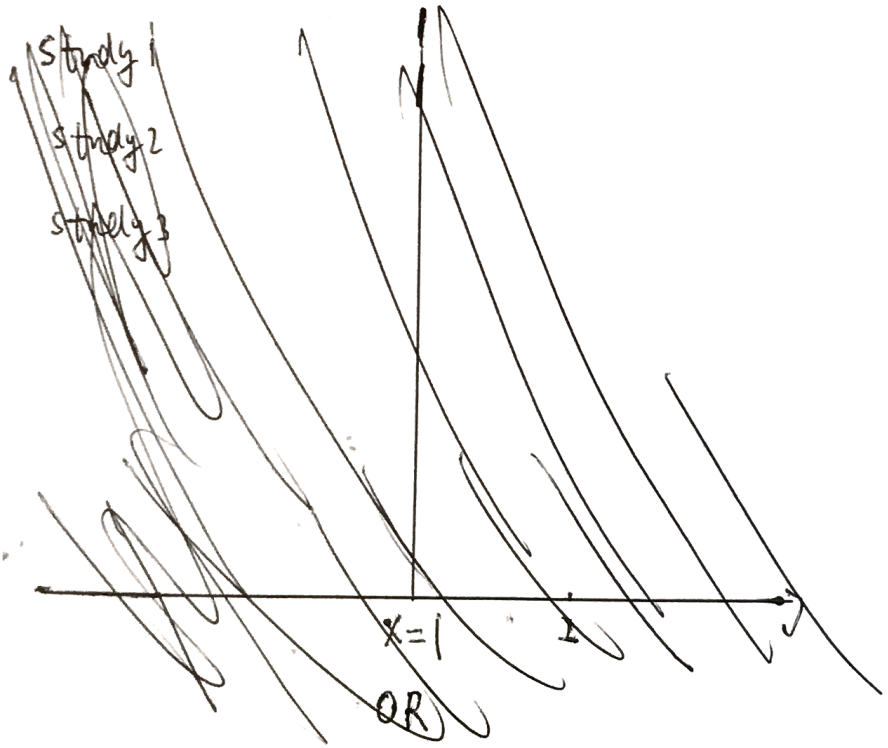
(1.029, 1.193)

3.5/4

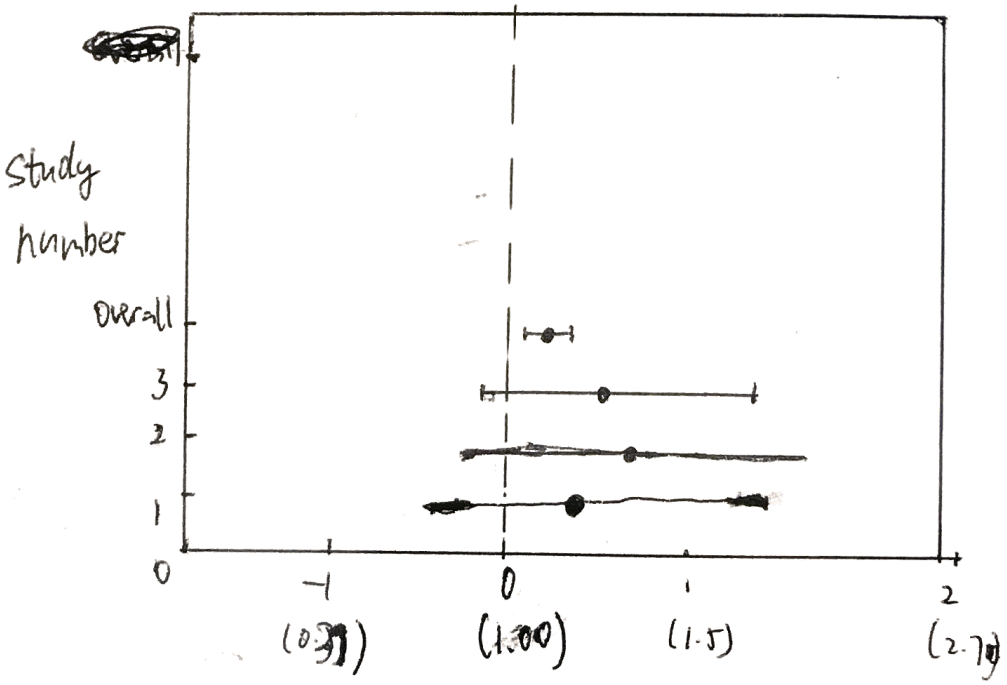
256155 1
P3

(viii) If a person's unhealthy diet increase by 1 unit, the odds of developing CHD increases by 11.~~6~~%, with 95~~%~~ percent confidence interval of increasing ranging from 2.9% to 19.3%. Since 95% of Mantel-Haenzel ~~Z~~ethal does not contain 1, the effect is statically significant. 1/2

(ix)



Graphic display of results of meta analysis



✓

5/5

2561551

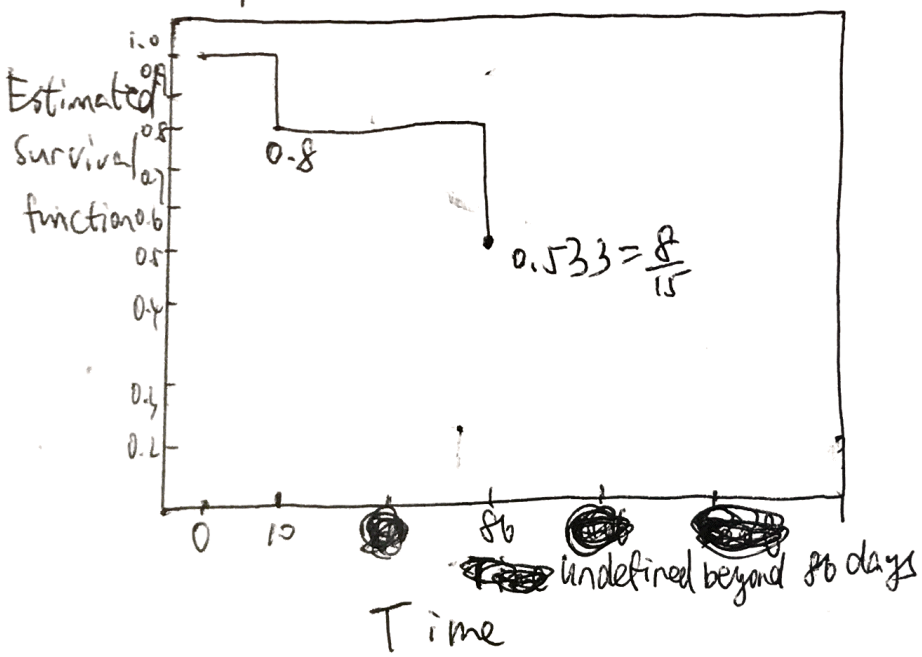
P4

2. (i)

Event time: $t_{(i)}$	At risk at time $t_{(i)}$: $r_{(i)}$	surviving beyond $t_{(i)}$: $f_{(i)}$	Estimated survival function at $t_{(i)}$: $\hat{S}(t_{(i)})$
0	5	5	$\frac{5}{5}=1$
10	5	4	$1\times\frac{4}{5}=0.8$
86	3	2	$0.8\times\frac{2}{3}=0.533=\frac{8}{15}$

√ 3/3

(ii) Kaplan-Meier Estimator illustration



2/2

(iii) $P(\text{survive at least 30 days}) = 0.8$ //

1/1

(iv) $S(t)=\exp(-\int_0^t h(u)du)=\exp(-H(t)) \Leftrightarrow H(t)=-\ln(S(t))$

at time point $(t=0)$: $H(0)=-\ln 1=0$

at time pt $(t=10)$ $H(10)=-\ln(0.8)\approx 0.223$

$t=86$, $H(86)=-\ln(\frac{8}{15})\approx 0.629$ //

√ 2/2

256/554

P5

(v.) $\hat{H}_{NA}(t) = \sum_{j=1}^{i} \frac{d_j}{r_j}$, where $d_j = r_j - s_j$

$\Rightarrow$ at $t=0$. $\hat{H}_{NA}(0) = \frac{5-5}{5} = 0$

at $t=10$ $\hat{H}_{NA}(10) = \frac{5-4}{5} + 0 = \frac{1}{5}$

at $t=86$, $\hat{H}_{NA}(86) = 0 + \frac{1}{5} + \frac{3-2}{3} = \frac{8}{15}$ ✓ 2/2

(vi) parametric analysis means fitting specific families of distribution to the survival data, for example, log-normal, gamma etc.

Nelson and Aalen cumulative hazard estimator is a non-parametric estimator. ✓ 2/2

(b)

(i) H_0: The survival functions for Group 1 and Group 2 are the same

H_1: The survival functions for Group 1 and Group 2 are different 2/2

(ii)

$t_{(i)}$	d_{1i}	d_{2i}	r_{1i}	r_{2i}	$E_{1i} = \frac{r_{1i}}{r_i} d_i$	$E_{2i} = \frac{r_{2i}}{r_i} d_i$
42	1	0	4	4	$\frac{4}{8} \times 1 = \frac{1}{2}$	$\frac{4}{8} \times 1 = \frac{1}{2}$
227	1	0	2	4	$\frac{2}{6} \times 1 = \frac{1}{3}$	$\frac{4}{6} \times 1 = \frac{2}{3}$
1996	0	1	0	2	0	$\frac{2}{2} \times 1 = 1$
2878	0	1	0	1	0	$\frac{1}{1} \times 1 = 1$
	$O_1 = 2$	$O_2 = 2$			$E_1 = \frac{5}{6}$	$E_2 = \frac{19}{6}$

$$X^2_{LR} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} = 2.063 < X^2_1 = 3.84$$ 5/5

256/571

P6

(iii) Since $X^2_{LR} = 2.063 < X^2 = 3.84$

⇒ we fail to reject H_0 and conclude that the survival function for Group 1 and Group 2 are the same. //

2/2

(c)

(i) hazard rate for age: $\exp(\hat{\beta}_1) = 1.03$

Interpretation: Increase the age by 1 unit, the hazard rate increase by 3%

2/3

(ii) ~~$h_0(t)$ is a arbitary baseline hazard function, evaluated when all the covariates are 0.~~

~~z_i is the covariate vector =~~

$h_0(t)$ is an arbitary baseline hazard function, when age and ts are all at default value 0.

z_i is the covariate vector = $\begin{pmatrix} \text{age (in years)} \\ \text{ts (continuous)} \end{pmatrix}$ ✓

β is the set of parameter: $\beta^T = (\beta_1: \text{age}, \beta_2: \text{ts})$ 1

pf Let $z_1 = \begin{pmatrix} 12 \\ 1.46 \end{pmatrix}$ $z_2 = \begin{pmatrix} 13 \\ 1.48 \end{pmatrix}$ 2/0

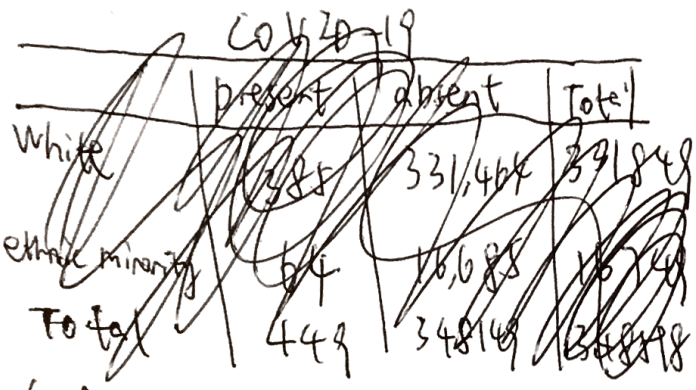
2.5/3

$$\frac{h(t, z_1)}{h(t, z_2)} = \frac{h_0(t)\exp(z_1^T\beta)}{h_0(t)\cdot\exp(z_2^T\beta)} = \exp\left((z_1 - z_2)^T\beta\right)$$ irrelevant of time ✓ 4

2356155

P7

3. (a)(i)



Covid-19	present	absent	total
ethnic minority	64 A	16685 B	16749
White	385 C	331464 D	331849
Total	449	348149	348598

2/2

(ii)

$P_1 = P(\text{Disease} \mid \text{White}) \approx 0.00116$

$P_2 = P(\text{Disease} \mid \text{minority}) \approx 0.00382$

AR = 2.648 per 1000 ppl

Interpretation: the absolute change in risk is an increase of 2.648 people per 1000 people if you have the risk factor ethnic minority.

2/2

(iii) $RR = \frac{\frac{64}{64+16685}}{\frac{385}{331464+385}} = 3.294$

$\text{Var}(\ln RR) = \frac{1}{A} - \frac{1}{A+B} + \frac{1}{C} - \frac{1}{C+D} = 0.018$

95% CI in lg scale: (0.929, 1.455)

95% of RR: (2.532, 4.284)

575

There is almost 3-fold incressed risk if you are ethic minorities, and the effect is statically significant as the 95% CI interval does not include 1.

25615H

P8

(iv) $$OR = \frac{\frac{64}{1685}}{\frac{385}{351464}} \approx 3.302$$

⇒ there is almost 3-folds increased risk of getting COVID-19 if you are ethnic minorities compared to white subjects per 1000 ppl. In 1000 people, the number of people getting COVID of ethnic minority is almost 3 times higher than white subjects. 2/2

(v.) No, they are not. ~~OR measures the ratio difference while~~

$$RR = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}, \quad OR = \frac{\frac{A}{B}}{\frac{C}{D}}$$

0/2

odds ratio measure how risk effect after the outcome, whereas relative risk measure the whole group, meaning that OR is not stable as RR.

(vi) ~~Poisson, relative risks, logistic regression~~

Study A measures relative risks, study B measures odds ratio.

Because in possion regression relative risk is natural in possion mean, while odds ratio fits well is logistic regression. 2/2

2561551

P9

(b) (i) sensitivity = $\frac{\text{number of diseased people screen positive}}{\text{Total number of diseased ppl}}$

If a screen test has low sensitivity, which means given the total number of diseased ppl, the test cannot identify the disease well, it only identify a few people who have disease, there are many people with disease, and the screen test cannot identify.

V1

(ii) sensitivity = $\frac{214}{372} = 0.575$

specificity = $\frac{5291}{5312} = 0.996$

(FNR) false negative rate = $1 - 0.575 = 0.425$

(FPR) false positive rate = 0.004

A sensitivity of 0.575 means 57.5% of diseased people are correctly identified as having COVID-19.

A false negative rate of 0.425 means 42.5% of diseased people are not correctly identified as having COVID-19.

A specificity of 0.996 means 99.6% of healthy people are correctly identified as not having COVID-19.

A false positive rate of 0.004 means 0.4% of healthy people are not correctly identified as having COVID-19.

5/5

P10

I recommend strategy A,

Consider two tests in row in plan B

	COV2D-19			
	Y		N	
positive flow test	214	123 correct	21	20.9 correct
		91 wrong		0.1 wrong
negative flow	158	91 correct	5291	5273 correct
		67 wrong		21 wrong

A not B

Sensitivity (B) = $\frac{123}{372} = 0.331 = P(\text{two screen positive} \mid \text{diseased})$

Specificity (B) = $\frac{5273}{5312} = 0.99 = P(\text{two screen negative} \mid \text{absent})$

For plan A.

since two event independent, sensitivity (A) = 0.575

specificity (A) = 0.996

⇒ choose plan A.

X 0/9