

Kontrol for én variabel i den statistisk analyse

Morten Frydenberg
Institut for Biostatistik®

Hvorfor er der et problem - et behov ?

Ideen bag den stratificerede analyse.

Ideen bag det korrigerede estimat.

Effektmodifikation - et eksempel.

Stratificerede analyser i STATA - ikke i hånden.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

1

Eksempel 1: Sammenligning af to behandlinger

Vi har samlet data med information om hvilken **behandling** 620 patienter fik og hvorvidt de **døde** indenfor 2 år.

Behandling	Død	I live	Total	% døde
A	200	100	300	67
B	160	160	320	50
Total	360	260	620	58

Som mål for forskellen mellem de to behandlinger kan vi estimere en **relativ risiko**: $\widehat{RR} = 1.33(1.16;1.53)$

Konklusion:

Risikoen for at dø er et sted mellem **16%** og **53%** større ved behandling A!

Behandling B er langt at foretrække!!!

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

2

Sammenligning af to behandlinger

Men !!!!!

- Mænd har en **dårligere prognose** end kvinder.
- Mænd fik oftere **behandling A** end kvinder gjorde.

Dvs. de, der fik behandling A, havde a priori en dårligere prognose (de var som regel mænd) end de, der fik behandling B (de var som regel kvinder).

Det vil således ikke være fair at sammenligne de to behandlinger uden at tage hensyn til, om det var mænd eller kvinder, man behandlede.



Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

3

Sammenligning af to behandlinger

Data delt op på køn (**stratificeret** efter køn):

Mænd	Behandling	Død	I live	Total	% døde
	A	180	60	240	75
	B	42	8	50	84
	Total	222	68	290	77

$$\widehat{RR}_M = 0.89(0.78;1.03)$$

Kvinder

	Behandling	Død	I live	Total	% døde
	A	20	40	60	33
	B	118	152	270	44
	Total	138	192	330	42

$$\widehat{RR}_K = 0.76(0.52;1.12)$$

Konklusion:

Både for mænd og kvinder tyder data på, at behandling **A** er den **bedste**!!!

Men for ingen af kønnene er vi helt sikre !

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

4

Sammenligning af to behandlinger

Mænd: $\widehat{RR}_M = 0.89(0.78;1.03)$

Kvinder: $\widehat{RR}_K = 0.76(0.52;1.12)$

Det tyder på, at den **relative risiko** godt kan antages at **være ens** for de to køn - forskellen mellem de to behandlinger ens for mænd og kvinder.

Det ville være rart med **ét estimat** for den relative risiko.

Et sådan estimat bør **ligge mellem de to estimater** vi fik for hver køn.

Det bør være et **passende gennemsnit** af **0.89** og **0.76**.

Det estimat ville man kalde: Det **korrigerede estimat** eller det **kombinerede estimat**.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

5

Sammenligning af to behandlinger

Mænd: $\widehat{RR}_M = 0.89(0.78;1.03)$

Kvinder: $\widehat{RR}_K = 0.76(0.52;1.12)$

Et vægtet gennemsnit: $\widehat{RR}_{Combined} = 0.84(0.71;0.99)$

Bemærk:

- Det **korrigerede estimat** ligger mellem de to kønsspecifikke estimater.
- **Sikkerhedsintervallet** for det korrigerede estimat er **smallere** end de to kønsspecifikke sikkerhedsintervaller.

Med andre ord: Et fælles estimat, der oven i købet er mere præcist end de to vi havde før!!!!

Konklusion: Behandling A bør foretrækkes!!

Der er et sted mellem **1%** og **29%** lavere dødelighed.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

6

Stratificeret analyse i STATA: Relativ Risiko

Data:

Køn	Behandling	Status	Antal
Mand	A	I live	60
Mand	A	Død	180
Mand	B	I live	8
Mand	B	Død	42
Kvinde	A	I live	40
Kvinde	A	Død	20
Kvinde	B	I live	152
Kvinde	B	Død	118

Kommando: `cs Status Behandling [fw=Antal], by (Køn)`

Output:

Køn	RR	[95% Conf. Interval]	M-H Weight
Mand	.8928571	.7751939 1.02838	34.75862
Kvinde	.7627119	.5202396 1.118195	21.45455
Crude	1.333333	1.164172 1.527075	
M-H combined	.8431854	.7146134 .9948898	

Test of homogeneity (M-H) $\chi^2(1) = 0.894$ $Pr>\chi^2 = 0.3443$

Morten Frydenberg MPH, Introduktionsmodul - Uge 3 - Dag 4

7

Stratificeret analyse i STATA: Relativ Risiko

Output:

Køn	RR	[95% Conf. Interval]	M-H Weight
Mand	.8928571	.7751939 1.02838	34.75862
Kvinde	.7627119	.5202396 1.118195	21.45455
Crude	1.333333	1.164172 1.527075	
M-H combined	.8431854	.7146134 .9948898	

Test of homogeneity (M-H) $\chi^2(1) = 0.894$ $Pr>\chi^2 = 0.3443$

Estimatet for mænd

Det ukorrigerede estimat

Estimatet for kvinder

Det korrigerede estimat

Vægte: Estimatet for mændene vægtes højere end det for kvinderne, da dette estimat er mindre usikkert.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

8

Stratificerede analyser / korrektion i analysefasen Generelt

Antag at vi er interesseret i at beskrive sammenhængen mellem en "eksponering" og et "outcome".

Men vi ønsker at korrigere for en tredje faktor, fx køn eller alder.

En begrundelse for et sådant ønske kunne være at den faktor er en **confounder**, dvs. en faktor der:

- er en **selvstændig risikofaktor**
- er **associeret med eksponering**
- **ikke** er et led i en konsekvens af eksponeringen
fx: **genindlæggelse er et led i årsagskæden**

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

9

Stratificerede analyser / korrektion i analysefasen Generelt

I den **stratificerede** analyse:

1. Opdeler man data i efter den tredje faktor - i **strata**.
fx (korrektion for køn): i mand og kvinder
fx (korrektion for alder): i 20-40, 41-50, 51-65 årige
2. I hvert stratum findes et **stratum specifikt estimat** for sammenhængen mellem eksponering og outcome.
3. Det vurderes om det er rimeligt antage at sammenhængen mellem eksponeringen og outcome kan **antages at være ens** for alle strata. **Homogenitet**.
fx: Er effekten af behandling A i forhold til B den samme for de to køn?
4. Hvis dette er rimeligt, så beregnes et **fælles/korrigeret** estimat, hvor de strata med mest information vægtes mest.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

10

Stratificerede analyser / korrektion i analysefasen Generelt

Nogle kommentarer:

ad.1 Hvilke strata, man skal inddele i afhænger af konteksten - strata skal være **så små**, at man kan antage at 'effekten' af eksponeringen er den samme indenfor hvert stratum.

Men **ikke** så små at de strata specifikke estimater bliver for usikre.

ad.3 Hvis association mellem eksponering og outcome ikke kan antages at være ens alle strata, så vil det **ikke give mening** at beregne et fælles/korrigeret estimat.

ad.4 **Beregning** af det fælles/korrigerede estimat er noget omstændigt og laves derfor **vha. pc**.

Der findes **forskellige metoder** - vi vil **ikke se** på dette her! - måske senere.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

11

Et case-reference studie af larynx cancer.

Et **case-reference** studie af sammenhængen mellem alkohol og larynx cancer.

310 cases og 1030 kontroller.

Hvad ved vi **a priori** om problemstillingen:

- Et **stort tobaksforbrug** er en kendt risiko faktor for larynx cancer.
- Tobaksforbrug og alkoholforbrug er **associeret**.



Vi **ønsker** vurdere alkohols betydning **renset for** tobaksforbrug.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

12

Et case-reference studie af larynx cancer.

Da de, der drikker meget, nok ofte ryger meget
og de, der intet drikker, ofte også er ikke-rygere,
vil vi komme til at **overvurdere** betydningen af det at
drikke, hvis vi ikke **korrigerer** for rygevaner.

Et **ukorrigerede** estimat vil således være for stort!

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

13

Stratificeret analyse i STATA: Odds Ratio

Data:

røg	alkohol	cancer	antal
0	0	0	450
1	0	0	240
0	1	0	110
0	1	1	15
0	0	1	55
1	1	1	116
1	0	1	124
1	1	0	230

0: Nej
1: Ja

Kommando: **cc cancer alkohol [fw=antal], by (røg)**

Output:

røg	OR	[95% Conf. Interval]	M-H Weight
0	1.115702	.5636589 2.094472	9.603175 (exact)
1	.9761571	.706377 1.348628	40.16901 (exact)
Crude	1.485212	1.134222 1.941398	(exact)
M-H combined	1.003081	.7599496 1.323999	

Test of homogeneity (M-H) $\chi^2(1) = 0.15$ $Pr>\chi^2 = 0.7013$
Test that combined OR = 1:
Mantel-Haenszel $\chi^2(1) = 0.00$
 $Pr>\chi^2 = 0.9826$

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

14

Stratificeret analyse i STATA: Odds Ratio

Output:

røg	OR	[95% Conf. Interval]	M-H Weight
0	1.115702	.5636589 2.094472	9.603175 (exact)
1	.9761571	.706377 1.348628	40.16901 (exact)
Crude	1.485212	1.134222 1.941398	(exact)
M-H combined	1.003081	.7599496 1.323999	

Vægte

Estimatet for ikke-rygere Det ukorrigerede estimat
Estimatet for rygere Det korrigerede estimat

Vægte: Estimatet for rygere **vægtes højere** end det for ikke rygere, da dette estimat er **mindre usikkert**.

Bemærk: De to **strata-specifikke** estimater er rimelig tæt på hinanden - det er ikke urimeligt at antage, at alkohol betyder det samme for rygere og ikke-rygere.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

15

Stratificeret analyse i STATA: Odds Ratio

røg	OR	[95% Conf. Interval]	M-H Weight
0	1.115702	.5636589 2.094472	9.603175 (exact)
1	.9761571	.706377 1.348628	40.16901 (exact)
Crude	1.485212	1.134222 1.941398	(exact)
M-H combined	1.003081	.7599496 1.323999	

Konklusioner:

1. Det **ukorrigerede** estimat er for stort!
2. Effekten af alkohol kan antages at være ens i de to strata! **Homogenitet!** - Lidt mere om det senere!
3. Når vi **korrigere** for rygning, så er der ikke den store effekt af alkohol **OR: 1.00(0.76;1.32)**.
4. **Hypotesen** om ingen effekt af alkohol, når vi korrigerer for rygning **kan accepteres**! $p=98\%$

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

16

Stratificeret analyse i STATA: Odds Ratio

Output:

røg	OR	[95% Conf. Interval]	M-H Weight
0	1.115702	.5636589 2.094472	9.603175 (exact)
1	.9761571	.706377 1.348628	40.16901 (exact)
Crude	1.485212	1.134222 1.941398	(exact)
M-H combined	1.003081	.7599496 1.323999	

Test of homogeneity (M-H) $\chi^2(1) = 0.15$ $Pr>\chi^2 = 0.7013$

Test that combined OR = 1:
Mantel-Haenszel $\chi^2(1) = 0.00$
 $Pr>\chi^2 = 0.9826$

Test for hypotesen om ingen forskel mellem strata

Test for hypotesen om ingen **OR=1** korrigeret for rygning

Denne stratificerede analyse, med odds ratio, som associationsmål kaldes en **Mantel-Haenszel analyse**.

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

17

Dødelighed i Danmark og Grønland.

6 aldersgrupper
Antal døde
Risikotid= population*1 år

agegroup	deaths	population	country
0	26	429	1
1	4	2044	1
2	11	7194	1
3	37	13572	1
4	35	2949	1
5	47	640	1
0	434	35625	0
1	101	149186	0
2	175	401597	0
3	1494	1076842	0
4	6166	552133	0
5	19204	288834	0

Grønland

Danmark

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

18

Dødelighed i Danmark og Grønland.

Analyse uden hensyntagen til alder:

Kommando: `ir deaths country population`

Output:

	Country		
	Exposed	Unexposed	Total
deaths	160	27574	27734
population	26828	2504217	2531045
Incidence Rate	.0059639	.011011	.0109575
	Point estimate		[95% Conf. Interval]
Inc. rate diff.	-.0050471	-.0059803	-.0041139
Inc. rate ratio	.5416314	.4607499	.6326697 (exact)
Prev. frac. ex.	.4583686	.3673303	.5392501 (exact)
Prev. frac. pop	.0048585		
	(midp) Pr(k<=160) =		0.0000 (exact)
	(midp) 2*Pr(k<=160) =		0.0000 (exact)

Dødeligheden er mindst i Grønland!!!

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

19

Dødelighed i Danmark og Grønland.

Analyse stratificeret for alder:

Kommando: `ir deaths country population ,by(agegroup)`

Output:

agegroup	IRR	[95% Conf. Interval]		M-H Weight	
0	4.974864	3.213031	7.390438	5.164087	(exact)
1	2.890585	.7726793	7.632454	1.3651	(exact)
2	3.50893	1.71992	6.434533	3.079691	(exact)
3	1.964985	1.378044	2.721128	18.59529	(exact)
4	1.062756	.7395521	1.479674	32.75828	(exact)
5	1.104522	.8112772	1.469373	42.45825	(exact)
Crude	.5416314	.4607499	.6326697		(exact)
M-H combined	1.534437	1.312696	1.793636		

Test of homogeneity (M-H) $\chi^2(5) = 54.53$ $P > \chi^2 = 0.0000$ **Alle de alderspecifikke incidens rate ratioer er større end 1.****De falder dog med alderen! Ikke homogenitet!****Et korrigeret estimat giver ingen mening!**

Morten Frydenberg

MPH, Introduktionsmodul - Uge 3 - Dag 4

20