NTNU
Det skapende universitet

Meta-analyse:
Modeller og metoder

Stian Lydersen

KLMED8006 Anvendt medisinsk statistikk
St2303 medisinsk statistikk
28 April 2010

Contents

Introduction to Bayesian methods

Partly based on
Everitt, B. S.: Modern Medical Statistics.
Arnold Publishers, 2003. Section 8.2

Frequentist vs Bayesian statistics.

The probability distribution $P(D|\theta)$ of our data D depends on some parameter(s) θ.

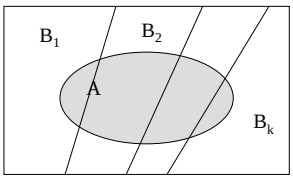
Example:

$$\underbrace{X_1,...,X_n}_{\text{Data } D} \sim N(\underbrace{\mu,\sigma^2}_{\text{Paramter } \theta})$$

The frequentist regards θ as an unknown constant

The Bayesian regards θ as an (unobserved) random variable from a probability distribution, *prior distribution*, $P(\theta)$.

Bayes rule (Rosner, eqn 3.10)



$$P(B_j \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$$

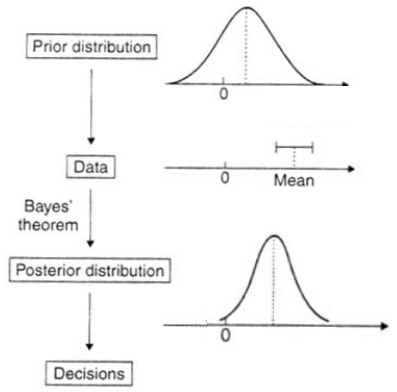
Following Bayes' theorem, the probability distribution of the parameter given the data, *posterior distribution*, is

$$P(\theta \mid D) = \frac{P(\theta)P(D \mid \theta)}{\int P(\theta)P(D \mid \theta)d\theta}$$

The denominator is a constant (does not depend on θ), so

$$P(\theta \mid D) \propto P(\theta)P(D \mid \theta)$$

posterior distribution $\propto$ prior distribution × likelihood



Bayesian estimate: $\hat{\theta}_B = E(\theta \mid D)$

(1-α) Bayesian confidence interval (credibility interval):
The α/2 quantiles (θ_l , θ_h) of the posterior distribution

Interpretation (!):

$P(\theta_l < \theta < \theta_h) = 1-\alpha$

Example:
Normal distribution.
$X_1,...,X_n \sim N(\mu,\sigma^2)$

Prior distribution:
$\mu \sim N(\nu,\tau^2)$

Posterior distribution:
$\mu \mid (X_1,...,X_n) \sim N(B\nu + (1-B)\bar{X}, (1-B)\sigma^2/n)$

where $B = \dfrac{\sigma^2/n}{\sigma^2/n+\tau^2}$

Bayes estimate (posterior mean)
$\hat{\mu}_B = B\nu + (1-B)\bar{X}$

A weighted average between the prior mean ν and $\bar{X}$

If the prior variance τ^2 is large,
we have a *vague* or *uninformative* prior,
and $\hat{\mu}_B \approx \bar{X}$

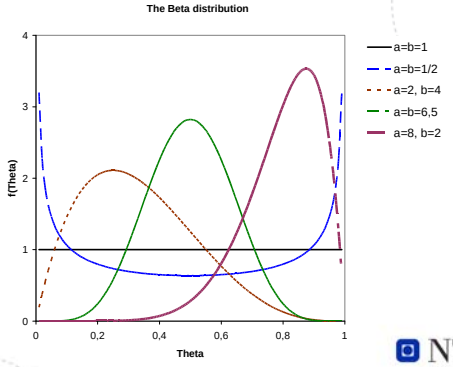
Example:
Binomial distribution:
number of events in n trials: $X \sim bin(n,\theta)$

Prior distribution:
$\theta \sim Beta(a,b)$

$f(\theta) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ for $0<\theta<1$

a>0, b>0

$E(\theta) = \dfrac{a}{a+b}$



13

Posterior distribution:

$\theta \mid X \sim Beta(a+x, b+(n-x))$

Posterior mean:

$$E(\theta \mid x) = \frac{a+x}{a+b+n}$$

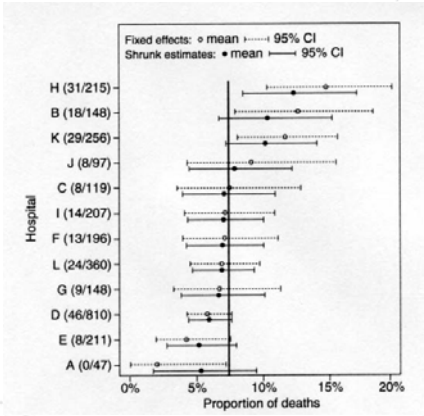
a = “pseudo number of events”
a+b=”pseudo number of trials”

14

Table 8.1 Mortality rates in 12 hospitals performing cardiac surgery in babies

	Hospital											
	A	B	C	D	E	F	G	H	I	J	K	L
No. of operations (n)	47	148	119	810	211	196	148	215	207	97	256	360
No. of deaths (r)	0	18	8	46	8	13	9	31	14	8	29	24

15



16

Meta Analysis. Models and Methods.

Mainly based on:

Normand, Sharon-Lise T: Tutorial in Biostatistics. Meta-Analysis: Formulating, Evaluating, Combining, and Reporting. *Statistics in Medicine*, **18**, 321-359 (1999)

17

Def.

Meta-Analysis may be broadly defined as the quantitative review and synthesis of the results of related but independent studies.

Meta-analyse er gjennomgang og sammenfatning av relaterte, uavhengige studier.

18

Examples:

- Randomized controlled trials of lidocaine vs placebo for patients with myocardial infarction
- LOS (length of stay) in hospital for stroke patients, specialist inpatient stroke care vs nonspecialist stroke care

19

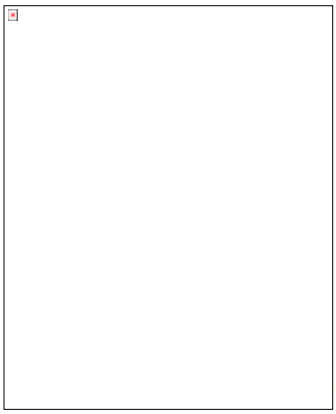
Table I. Prophylactic use of lidocaine after a heart attack: evaluating mortality from prophylactic use of lidocaine in acute myocardial infarction. Source: reference 1

Source	Number randomized		Number dead	
	Lidocaine	Control	Lidocaine	Control
1. Chopra *et al.*	39	43	2	1
2. Mogensen	44	44	4	4
3. Pitt *et al.*	107	110	6	4
4. Darby *et al.*	103	100	7	5
5. Bennett *et al*	110	106	7	3
6. O'Brian *et al.*	154	146	11	4
Total	557	549	37	21

NTNU Det skapende universitet

www.ntnu.no

20



NTNU Det skapende universitet

www.ntnu.no

21

Table II. Specialist care for stroke patients from nine studies: comparing specialist multidisciplinary team care for managing stroke inpatients with routine management in general medical wards. Source: reference 2

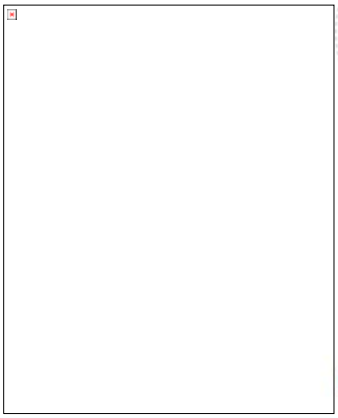
Source	Specialist care			Routine management		
	N	Mean LOS	SD	*N*	Mean LOS	SD
1. Edinburgh	155	55·0	47·0	156	75·0	64·0
2. Orpington-Mild	31	27·0	7·0	32	29·0	4·0
3. Orpington-Moderate	75	64·0	17·0	71	119·0	29·0
4. Orpington-Severe	18	66·0	20·0	18	137·0	48·0
5. Montreal-Home	8	14·0	8·0	13	18·0	11·0
6. Montreal-Transfer	57	19·0	7·0	52	18·0	4·0
7. Newcastle 1993	34	52·0	45·0	33	41·0	34·0
8. Umea 1985	110	21·0	16·0	183	31·0	27·0
9. Uppsala 1982	60	30·0	27·0	52	23·0	20·0
Total	548			610		

LOS = length of stay measured in days; SD = standard deviation.

NTNU Det skapende universitet

www.ntnu.no

22



NTNU Det skapende universitet

www.ntnu.no

23

Fixed effect vs random effect models

- Fixed effects: The studies have identical characteristics and study effects.
- Random effects: The studies may have different effects and different characteristics
- Debate as to the choice of appropriate model
- *Always* reasonable to assume some between-study variation and few reasons to believe it is zero.

NTNU Det skapende universitet

www.ntnu.no

24

Combining studies: Different classes of outcome (study summary)

- Discrete outcome such as difference in proportions
- Continuous outcome such as means
- Test statistics

(not exhaustive list)

NTNU Det skapende universitet

www.ntnu.no

25

Binary outcome:

		outcome: Yes	outcome: No	
Group	T	a_i	b_i	$a+b=n_{Ti}$
	C	c_i	d_i	$c+d=n_{Ci}$

Study number i: $\hat{p}_{Ti} = \frac{a_i}{n_{Ti}}, \quad \hat{p}_{Ci} = \frac{c_i}{n_{Ci}}$

26

Risk differences, relative risks, odds ratios
Etimators and confidence intervals as described in Rosner (2005), Section 13.3

	Estimator	Standard deviation
Risk difference	$d_i = \hat{p}_{Ti} - \hat{p}_{Ci}$	$s_{d_i} = \sqrt{\frac{p_{Ti}(1-p_{Ti})}{n_{Ti}} + \frac{p_{Ci}(1-p_{Ci})}{n_{Ci}}}$
Relative risk (Risk ratio)	$r_i = \hat{p}_{Ti} / \hat{p}_{Ci}$	$s_{Log(r_i)} = \sqrt{\frac{1-p_{Ti}}{n_{Ti}p_{Ti}} + \frac{1-p_{Ci}}{n_{Ci}p_{Ci}}}$
Odds ratio	$\omega_i = \frac{\hat{p}_{Ti}/(1-\hat{p}_{Ti})}{\hat{p}_{Ci}/(1-\hat{p}_{Ci})}$	$s_{Log(\omega_i)} = \sqrt{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}}$

27

Continuous outcome:

Difference in means from study number i:

$Y_i = \bar{x}_{Ti} - \bar{x}_{Ci}$

with standard deviation s_i calculated as

$s_i^2 = s_{pi}^2 \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} \right)$

where $s_{pi}^2 = \frac{(n_{Ti}-1)s_{Ti}^2 + (n_{Ci}-1)s_{Ci}^2}{n_{Ti}+n_{Ci}-2}$

28

Possible effect size: The standardized mean difference

$\delta_i = \frac{\mu_i^T - \mu_i^C}{\sigma_i}$

when

$Y_{ij}^T \sim N(\mu_i^T, \sigma_i^2); \quad j = 1,2,...,n_{Ti}$

$Y_{ij}^C \sim N(\mu_i^C, \sigma_i^2); \quad j = 1,2,...,n_{Ci}$

(Normand, 1999, states the above without subscript i)

29

Estimator for δ (denoted Hedges' g):

$h_i = \frac{\bar{Y}_i^T - \bar{Y}_i^C}{s_i},$

$\hat{Var}(h_i) = \left(\frac{1}{n_{Ti}} + \frac{1}{n_{Ci}} \right) + \frac{\hat{\delta}_i^2}{2(n_{Ti}+n_{Ci})}$

where $\hat{\delta}_i^2$ is the sample estimate of δ_i^2.

30

p-values as outcome

Many methods exist for combining $p_1, p_2, \ldots, p_k$.
(Cooper and Hedges (1994) list 16 methods)

Much used (Darlington & Hayes, 2000):
The Stouffer and the Fisher method

Under H_0, under quite general conditions,
p_i is (approximately) uniformly distributed on (0,1)

31

The Stouffer (1949) method:

Compute the z-values corresponding to the p-values:

$z_i = \Phi^{-1}(p_i) \sim N(0,1)$ under H_0

Combined z-value: $z = \frac{z_1 + z_2 + \ldots + z_k}{\sqrt{k}} \sim N(0,1)$ under H_0

Combined p-value: $p = \Phi(z)$

Example (Darlington & Hayes, 2000):
p-values .159, .133, .111, .092
z-values -.999, -1.112. -1.221, -1.329
combined z=-2.330, combined p=0.010

32

The Fisher (1932) method:

A low value of $p_1 p_2 \cdot \ldots \cdot p_k$ or equivalently,
$\log(p_1 p_2 \cdot \ldots \cdot p_k) = \sum_{i=1}^{k} \log(p_i)$,
is taken as evidence against H_0

Under H_0, $-2\sum_{i=1}^{k} \log(p_i) \sim \chi^2_{2k}$

Example:
p-values .159, .133, .111, .092
χ^2 = 16.881, df= 2*4=8
Fisher p=0.031

33

Publication bias:

File-Drawer (Fail-Safe) numbers

Researchers may have unpublished, not significant results “in their file-drawers”

How many unknown studies ($N_{FS,\alpha}$) with average z=0 need to be added to the known N to make the outcome of Stouffer’s test not significant at level α? (Rosenthal, 1979)

34

Sum of $N_{FS,\alpha}$ terms

$$\left|\frac{\sum_{i=1}^{N} z_i + 0}{\sqrt{N_{FS,\alpha} + N}}\right| \le z_{\alpha/2} \quad \text{or} \quad N_{FS,\alpha} \ge \left(\frac{\sum_{i=1}^{N} z_i}{z_{\alpha/2}}\right)^2 - N$$

Example (cont’d)
N=4, combined p=0.010,
combined $z=z_{1\text{-}0.010}$ = -2.330, $\sum z_i = -2.330\sqrt{4} = -4.660$

$N_{FS,0.05} \ge \left(\frac{-4.660}{1.96}\right)^2 - 4 = 1.65$ dvs $N_{FS,0.05}$=2.

35

but:

Combining p-values gives little insight in effect size and its direction.

36

Sources of variation in meta-analysis

- Sampling error may vary between studies. Varying sample size.
- Study-level characteristics may vary (for example for-profit vs not-for-profit hostpitals)
- Inter-study variation (random effects model)

37

Y_i = summary statistic from study no i
(for example treatment effect)

Approximately normally distributed

Fixed effect model:

$Y_i \sim N(\theta, s_i^2)$ for $i = 1, 2, ..., k$

where θ is the mean treatment effect (same in all studies)

38

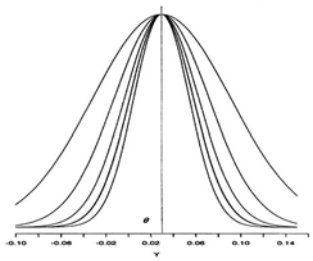


Fig. 3. Fixed effect model.

39

Random-effects model
Similar to:

- multi level model
- hierarchical model
- mixed effect model

$Y_i \mid \theta_i, s_i^2 \sim N(\theta_i, s_i^2)$

where θ_i is the study specific mean drawn from a superpopulation with hyperparameters θ and τ^2,

$\theta_i \mid \theta, \tau^2 \sim N(\theta, \tau^2)$

40

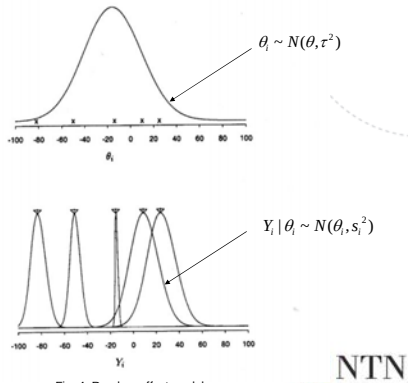


Fig. 4. Random effect model.

41

Unconditional distribution of Y_i:

$Y_i \sim N(\theta, s_i^2 + \tau^2)$

where s_i^2 and τ^2 are within-study and between study variations.

42

The distribution of the study-specific effect θ_i, conditional on the observed data and the hyperparameters, is

$\theta_i \mid (Y_1, ..., Y_k), \theta, \tau^2 \sim N(B_i\theta + (1 - B_i)Y_i, s_i^2(1 - B_i))$

where $B_i = \frac{s_i^2}{s_i^2 + \tau^2}$ is the *shrinkage factor* for the i'th study

43

Fixed-effect model:

Maximum likelihood estimator for common mean if s_i^2 is known:

$$\hat{\theta}_{MLE} = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \quad \text{with} \quad W_i = \frac{1}{s_i^2},$$

$$\hat{\theta}_{MLE} \sim N\left(\theta, \left(\sum_{i=1}^{k} W_i\right)^{-1}\right)$$

44

Test for homogeneity of study means:

H_0: $\theta_1 = \theta_2 = ... = \theta_k$

H_1: At least two are different

Under H_0, for large sample sizes,

$$Q_W = \sum_{i=1}^{k} W_i (Y_i - \hat{\theta}_{MLE})^2 \sim \chi^2_{k-1}$$

45

Random-effects model:

Maximum likelihood estimator for common mean if τ^2 was known:

$$\hat{\theta}(\tau)_{MLE} = \frac{\sum_{i=1}^{k} W_i(\tau) Y_i}{\sum_{i=1}^{k} W_i(\tau)} \quad \text{with} \quad W_i(\tau) = \frac{1}{s_i^2 + \tau^2}$$

Usually τ^2 is unknown.
Two common estimation methods are

- REML (*Restricted Maximum Likelihood*)
- *Bayesian*

46

REML (Restricted Maximum Likelihood)

- Apply linear functions **K'y** such that **K'y** contains none of the fixed effects.
- Estimate the random effects (variance parameters) by applying ML to **K'y**.
- For fixed effects, REML = ML.

47

Example:
When $X_1, X_2, ..., X_n$ independent $N(\mu, \sigma^2)$

$$\hat{\sigma}^2_{REML} = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

48

REML or ML?

- (Diggle & al, 2002, p 69).
  - REML estimators should be less biased
- (McCulloch and Searle, 2001, p 177-178)
  - A growing preference for REML in mixed models
  - For balanced and normal data, REML solutions are minimal variance unbiased.
  - REML for unbalanced data sets yield no clean algebraic results
  - REML estimators seem to be less sensitive to outliers in the data

49

REML:
Estimates of θ and τ^2 may be found as solutions to

$$\hat{\tau}_R^2 = \frac{\sum_{i=1}^{k} w_i^2(\hat{\tau}_R)\left(\frac{k}{k-1}(Y_i-\hat{\theta}_R)^2 - s_i^2\right)}{\sum_{i=1}^{k} w_i^2(\hat{\tau}_R)}$$

$$\hat{\theta}_R = \frac{\sum_{i=1}^{k} w_i(\hat{\tau}_R)Y_i}{\sum_{i=1}^{k} w_i(\hat{\tau}_R)} \quad \text{with} \quad w_i(\hat{\tau}_R) = \frac{1}{s_i^2 + \hat{\tau}_R^2}$$

50

Estimator for θ_i (empirical Bayes):

$$\hat{\theta}_i^R = \hat{B}_i^R\hat{\theta}_R + (1-\hat{B}_i^R)Y_i \quad \text{where} \quad \hat{B}_i^R = \frac{s_i^2}{s_i^2 + \hat{\tau}_R^2}$$

51

Full Bayesian approach:

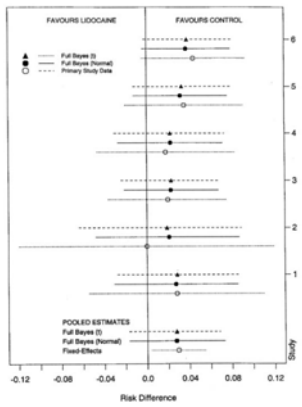
θ, τ^2 are regarded as random variables with one realization

For example

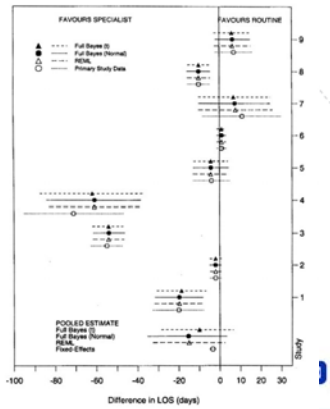
$\theta \sim N(0, a^2)$ and $\tau^{-2} \sim gamma(c, d)$

The hyperparameters (a,c,d) are specified a priori, not estimated from data

52



53



54

Mantel-Haenzel methods for 2x2 tables

Rosner, Section 13.5

55

K tables:

		outcome	
		Yes	No
Group	T	a_i	b_i
	C	c_i	d_i

NTNU Det skapende universitet

www.ntnu.no

56

Mantel-Haenzel test for conditional independence.

Under H_0, in table no i, conditional on the row and column sums, a_i is hypergeometric distributed (as in Fisher's exact test), with mean and variance

$$E_i = \frac{(a_i + b_i)(a_i + c_i)}{n_i}$$

$$V_i = \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2 (n_i - 1)}$$

NTNU Det skapende universitet

www.ntnu.no

57

The (Cochran-)Mantel-Haenzel test for conditional independence
Null hypotheses: The common OR=1:

$$\chi^2_{MH} = \frac{(O-E)^2}{V} \sim \chi^2_1 \quad \text{under } H_0$$

where

$$O = \sum_{i=1}^{k} O_i = \sum_{i=1}^{k} a_i$$

$$E = \sum_{i=1}^{k} E_i = \sum_{i=1}^{k} \frac{(a_i + b_i)(a_i + c_i)}{n_i}$$

$$V = \sum_{i=1}^{k} V_i = \sum_{i=1}^{k} \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{n_i^2 (n_i - 1)}$$

NTNU Det skapende universitet

www.ntnu.no

58

Rosner Eq 13.14 p 653 uses a continuity correction, as proposed by Mantel & Haenzel (1959).

$$\chi^2_{MH} = \frac{(|O-E| - 0.5)^2}{V}$$

This approximates an exact conditional test, but tends to be conservative (Agresti 2002)

NTNU Det skapende universitet

www.ntnu.no

59

The Mantel-Haenzel estimator for the common odds ratio:
Rosner, eqn 13.15 p 655

$$\hat{OR}_{MH} = \frac{\sum_{i=1}^{k} \frac{a_i d_i}{n_i}}{\sum_{i=1}^{k} \frac{b_i c_i}{n_i}}$$

NTNU Det skapende universitet

www.ntnu.no

60

Confidence interval for common OR (Rosner eqn 13.16)

$$\exp\left[\ln \hat{OR}_{MH} \pm z_{1-\alpha/2} \sqrt{Var(\ln \hat{OR}_{MH})}\right]$$

where

$$Var(\ln \hat{OR}_{MH}) = \frac{\sum_{i=1}^{k} P_i R_i}{2\left(\sum_{i=1}^{k} R_i\right)^2} + \frac{\sum_{i=1}^{k} (P_i S_i + Q_i R_i)}{2\left(\sum_{i=1}^{k} R_i\right)\left(\sum_{i=1}^{k} S_i\right)} + \frac{\sum_{i=1}^{k} Q_i S_i}{2\left(\sum_{i=1}^{k} S_i\right)^2}$$

and

$$P_i = \frac{a_i + d_i}{n_i}, \quad Q_i = \frac{b_i + c_i}{n_i}, \quad R_i = \frac{a_i d_i}{n_i}, \quad S_i = \frac{b_i c_i}{n_i}$$

NTNU Det skapende universitet

www.ntnu.no

61

Tests for homogeneity of odds ratios:

- Strata:
  - Woolf method, Rosner eqn 13.17,
  - Breslow-Day method (SPSS, StatXact)
- Meta-analysis:
  - Rosner eqn 13.40

NTNU Det skapende universitet

62

Example Doll and Hill (1950),
Rosner exercise 13.9 - 13.15

Men:	smoke	non-smoke	total
lung cancer	647	2	649
control	622	27	649
total	1269	29	1298

Women	smoke	non-smoke	total
lung cancer	41	19	60
control	28	32	60
total	69	51	120

NTNU Det skapende universitet

63

Men and women separately:

	estimate	95% c.i.	p-value
Men	14.1	3.3 to 59	2.7E-6
Women	2.47	1.2 to 5.2	0.017

exact StatXact

	estimate	95% c.i.	p-value
Men	14.1	3.5 to 122	1.3E-6
Women	2.47	1.1 to 5.6	0.026

Test for homogeneity of OR:
Breslow & Day statistic = 5.21, df=1, p=0.022

MH estimate for common OR: 4.52
95% c.i.: 2.42 to 8.47

NTNU Det skapende universitet

64

Logistic regression – an alternative to Mantel Haenzel methods.

- Strata (or study) as categorical covariate (or coded with k-1 indicator variables)
- Approximately same results as Mantel-Haenzel methods.

NTNU Det skapende universitet

65

References

- Cooper, H, Hedges, L. V. ”The Handbook of Research Synthesis.” Russell Sage Foundation, 1994.
- Darlington, R. B., Hayes, A. F.: Combining Independent p Values: Extensions of the Stouffer and the Binomial Models. *Psychological Methods*, 2000, Vol 5, No 4, 496 - 515
- Diggle, P. J., Heagerty, P., Liang, K-Y, Zeger, S. L.: Analysis of Longitudinal Data”2nd ed, Oxford University press, 2002.
- Everitt, B. S.: Modern Medical Statistics. Arnold Publishers, 2003.

NTNU Det skapende universitet

66

References

- McCulloch, C. E., Searle, S. R: ”Generalized, Linear and Mixed Models”. Wiley, 2001.
- Normand, Sharon-Lise T: Tutorial in Biostatistics. Meta-Analysis: Formulating, Evaluating, Combining, and Reporting. *Statistics in Medicine*, **18**, 321-359 (1999)
- Rosenthal, R. The “file-drawer problem” and tolerance for null results. *Psychological Bulletin*, **86**, 185-193 (1979)
- Rosner, B. Fundamentals of Biostatistics. 6th Ed. Thomson, 2005.

NTNU Det skapende universitet